The Phonological Encoding of Complex Morphosyntactic Structures in Native and Non-Native English Speakers

Hilary Suzanne Zinsmeyer Wynne

Lady Margaret Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2016

ABSTRACT

Theories of phonological word formation (e.g. Selkirk 1980, 1986; Nespor & Vogel 1986; Lahiri & Plank 2010) assume that prosodic units are not isomorphic with syntactic units. However, the prosodic status of compounds remain uncertain, at least in so far as language planning and phonological encoding is concerned. Theories are not transparent about the prosodic status of compounds: although a noun-noun compound in English consists of two lexical words (and therefore two prosodic words), it can also act as a single prosodic item by exhibiting main stress on the first unit and carrying inflection. Thus the question remains controversial - should these items be treated as a single prosodic unit, similar to a monomorphemic word, or as two distinct units for the purpose of post-lexical representation? Recursive word formation may suggest that compounds are a single unit. Psycholinguistic evidence measuring speech onset latency in native speakers of Dutch and Portuguese also shows compounds being treated as single prosodic units (Wheeldon & Lahiri 1997, 2002; Vigario, 2010). Although recent studies have produced evidence for the prosodification of compounds in native speakers, little is known about the process in non-native speakers. Our research questions are as follows: what is the post-lexical planning unit in English, and how do non-native fluent speakers of English plan these units for the purpose of phonological encoding?

To investigate our hypotheses, we focus on the phonological encoding of compounds with and without encliticisation, for native and non-native speakers of English. The latter group are of special interest to compare the encoding of compounds (with initial stress) to monomorphemic words varying in trochaic and iambic rhythm. Our initial hypotheses are that learning phrasal grouping, which includes prosodic word formation, is different from learning lexical word stress. To begin, we conducted two experiments containing sets of English stimuli containing noun-noun compounds, adjective-noun phrases, disyllabic initial-stressed words, and disyllabic final-stressed words (monosyllabic in Experiment 2). Experiment 1 looked at how compounds were planned in native English speakers, using both delayed and online task conditions. Experiment 2 presented the target stimuli in online task conditions, where speakers to begin speaking immediately after the question. This task used a monosyllabic word instead of a final-stressed word as one of the control conditions.

In Experiment 1, adjective-noun phrases elicited significantly longer naming latencies than all other conditions, while compounds showed no difference to either disyllabic word condition. In Experiment 2, adjective-noun phrases elicited significantly shorter naming latencies than both compounds and disyllabic words. Notably, the mean naming latency for the phrasal condition was statistically-similar to the monosyllabic condition.

In the delayed priming task, reaction times reflected the total number of prosodic units in the target sentence. In the online task, however, speech latencies only reflected the complexity of the first prosodic unit. Taken together, these results suggest that, despite containing two lexical and prosodic words, English compounds are planned as single prosodic units. The next set of experiments (Experiments 3 and 4) compared cliticisation in compounds with cliticisation in monomorphemic words. The results were as expected: in the delayed task, the adjective-noun phrase+ clitic condition elicited significantly longer latencies, while the online task results showed significantly shorter latencies for the phrasal condition to all other conditions.

Using the same stimuli, we ran similar experiments with fluent English speakers of Standard Colloquial Bengali in Calcutta. Despite the fact that there were many errors for monomorphemic words with second syllable stress, the Bengali speakers exhibited identical patterns of phonological encoding in both sets of experiments. For metrical stress, their own word-initial stress pattern inhibited their reactions. However, in terms of phonological encoding and phonological word formation, the results were entirely in line with the native English speakers. In Experiments 5 and 6, compounds were produced in exactly the same as monomorphemic words while adjective noun phrases were treated as two independent words. In Experiments 7 and 8, results were similar to the native English speakers: adjective-noun phrases continued to be planned as separate units.

In both delayed priming tasks, reaction times reflected the total number of prosodic units in the target sentence. In the online tasks, however, speech latencies only reflected the size of the first prosodic unit. Taken together, these results suggest that, despite containing two lexical and prosodic words, English compounds are planned as single prosodic units, exhibiting encliticisation and reaction times similar to those of monomorphemic words. These findings lend strong support to the claim that it is in fact the prosodic structure (not the lexical or morphosyntactic structure) of the utterance that is dictating the arrangement of prosodic frames during the post-lexical encoding stages of language production.

Word Count: 73,236

ACKNOWLEDGMENTS

This dissertation would never have existed without the guidance, support and friendship of a great number of people. First and foremost, I would like to thank my supervisor Aditi Lahiri without whom none of this would have been possible. She provided an endless source of encouragement and insight during both my M.Phil and D.Phil, and has become both a mentor and friend over the past five years.

I would also like to thank Linda Wheeldon, who has served as an invaluable source of feedback and support during this project. She spent long hours reading numerous drafts, sharing experimental files, and helping me with statistics over Skype. I am also indebted to my committee members, Professor E. Matthew Husband and Professor Paula Fikkert for their comments, suggestions, and careful reading of this thesis.

Endless amounts of gratitude and love also go, most deservingly, to my long-suffering parents and family. Thank you Mom, for the hours and hours of overseas phone calls, the cards, the unwavering support, and the dog photos. And thank you Dad, for never lowering your expectations of me; you knew I could do this, even when I didn't. Thank you to my best friend and sister Gwendolyn, for always being there and making me laugh.

As a famous band once quipped, what a long strange trip it's been. The road to this doctorate was neither straight nor simple, but the knowledge, friendship, and memories I've gained along the way have been priceless. My acknowledgments would be hollow without mention of Professor Michael Drout, who set me on this road and saw in me some inkling of academic capability all those years ago. Thank you, Michael, for pushing me to achieve what I never believed I could, and being a constant friend ever since.

On the topic of friends, my thanks go as well to all those who listened to my weekly dramas, in particular Dr. Anthony Adams, Dr. Sandra Kotzor, Alistair Hanson, Hugo Finley, Susana Hancock, Dr. Jeannique Darby, Jeanne Ryan, Dr. Steven Kaye, Dr. Holly Kennard, and Dr. Ellie Darlington. Thank you also to the rest of the LAB Lab, the Noons, GoonTube, CORC Tier 2 Women, and all the colleges I've rowed for during my time at Oxford.

My gratitude also goes to my subjects, both at Oxford and in India, for taking part in what could be rather long and arduous production tasks, and to all the assistants who helped me with the experimental runs in Kolkata.

> "I look at my yesterdays for months past, and find them as good a lot of yesterdays as anybody might want. I sit there in the firelight and see them all. The hours that made them were good, and so were the moments that made the hours. I have had responsibilities and work, dangers and pleasure, good friends, and a world without walls to live in."
> - *Beryl Markham, West with the Night*

Contents

List of Figures

List of Tables

1

Introduction

1.1 Motivation

The act of speaking is a complex process, requiring the simultaneous activation of dozens of muscles and multiple areas of the human brain (cf. Ackerman & Rieker, 2010). Models of speech production agree that there are a series of cognitive stages involved in the production of speech: each stage prepares corresponding representations, i.e. concepts are generated as representations at the semantic level, lemmas during lexicalisation, and phonological representations during post-lexical encoding. More recently, the processes dedicated to preparing the phonological features of speech have become the centre of focus.

The extent to which speech is planned prior to articulation is not entirely clear. Experimental evidence suggests that this process is affected by a number of things, including (but not limited to) the time a speaker has to prepare the utterance and the amount of cognitive load the situation requires. This thesis has two goals centred around this planning process: first, to investigate the planning of connected speech at the post-lexical encoding stage in English, and second, to investigate this process in non-native speakers of English. Both of these goals rely upon formal theories of phonological phrasing as well as evidence for the structure and behaviour of the post-lexical encoding processes in speech production.

1.2 Post-Lexical Planning

1.2.1 How We Plan

Contemporary models of speech (Levelt, 1989; et al., 1999; Roelofs, 1997), assume that each production level accepts a prepared representation as input and generates a new representation as its output. The post-lexical (also referred to as "phonological") encoding stage is the stage of speech production where phonemes, syllables, stress, and metrical structure are accessed in order to prepare the utterance for articulation. Phonological encoding occurs during a dedicated level, one that is actually made up of intermediate stages in which phonological features are activated according to the needs of the utterance. One of these sub-stages is dedicated to the preparation of prosodic frames, the frames which organise spoken language for articulation.

During this subprocess, prosodic structures are built by attaching segmental specifications to prosodic frames during syllabification. Information about the utterance's phonology is then combined with existing information in the morphosyntactic representation: syllabification, CV structure, moraic structure, stress, and intonation are among the features added at this stage.

1.2.2 What We Plan

Empirical research in multiple languages has indicated that the shape of these prosodic frames is motivated by a unit that is prosodically (rather than lexically) shaped: (Nespor & Vogel, 1986 and Vogel, 1991 for Italian; Booij & Rubach, 2008, Wheeldon & Lahiri, 1997, 2002, and Van Oostendorp, 2002 for Dutch; Drachman & Malikouti-Drachman, 1994, Nespor & Ralli, 1996, Revithiadou, 2008, and Nikolou, 2011 for Greek; Kabak & Vogel, 2001; Göksel, 2009 for Turkish, Nikolou, 2011, for Arabic, Hannahs, 1995, for French, Sandler, 1999 and Nespor & Sandler, 1999 for sign languages, and Vigário, 2003 for European Portuguese). Current models of speech production (Levelt, 1992; Roelofs, 1997) maintain that the output representation of phonological encoding is the Phonological Word representation, in which syllables and word stress are grouped in order to prepare for phonetic retrieval.

The Phonological Word is a unit of language defined in prosodic constituent theory as the domain where morphological and phonological structure interact (Selkirk, 1980b; Nespor and Vogel, 1986; Selkirk, 1996; Peperkamp, 1997). Minimally composed of at least one stressed foot, a Phonological Word can contain a lexical word plus any number of unstressed items. A lexical word is a well-formed semantic and syntactic unit that can stand on its own, be uttered in isolation, and even be considered a full utterance. For example, the lexical word *dog*, if uttered with a question intonation, can be construed as someone raising doubt about whether the animal in question is a dog or not. In contrast, *a dog* constitutes both a syntactic phrase and a single Phonological Word, where the determiner *a* attaches as a phonologically-weak element to *dog*: [ədɒg].

Theories of phonological phrasing (Selkirk, 1980a,b, 1986; Lahiri and Plank, 2010) assume that prosodic units are not isomorphic with syntactic units. Accordingly, the number of morphosyntactic words is often different to the number of Phonological Words in an utterance. For example, the phrase "understand it" (example 1.1 below) consists of two lexical (morphosyntactic) words, but only one Phonological Word[1] when spoken in normal or rushed speech:

(1.1) "understand it"
lexical words: $[\text{understand}]_{\text{LEX}}$ $[\text{it}]_{\text{LEX}}$
Phonological Words: $(\text{understand it})_{\omega}$

This mismatch points to a process in speech production where prosodically-shaped units are prepared. Early models of language production (Harley, 1984; Dell, 1986) were comprehensive in their attention to lexical processing; however, the output of this stage was typically treated as a lexical unit that eventually received phonological information. Later models (Levelt, 1989; Roelofs, 1999) began to treat the output of the lexicalisation stage as post-lexical representations: Levelt's original model (1989) described the output of grammatical encoding as a tree structure (called "the surface structure") in which lemmas are first retrieved from the mental lexicon as sublexical and subsyllabic units, then positioned in independent structures (such as word and syllable skeletons) during phonological encoding. The output of this process contained templates for syllable constituents and Phonological Word segments.

1.3 Current Approach

Nespor & Vogel (1986) write, "at the very core of linguistic theory is the claim that an appropriate theory should be able to account for the psychological reality of language" (p. xv). This thesis aims to do exactly that: while theoretical frameworks aim to explain **how** and **why** phonology and syntax interact differently, psycholinguistic models attempt to place **where** this happens in the cognitive framework of language generation. This thesis merges both the formal linguistic and psychological approaches to the syntax-phonology interface by eliciting evidence for the post-lexical encoding process. We present a series

[1]We endeavour to use the ω symbol to notate Phonological Word boundaries consistently throughout this work, unless a particular framework specifies a different practice (e.g. PWd). The same will apply to the notation for syllable (σ) and stress foot (Σ).

of experiments designed not only to examine the process that prepares prosodic structures in English, but also the shape of these structures themselves.

1.4 Research Questions

Results from interference tasks containing phonologically-related distractors have suggested that the scope of phonological encoding process is smaller than that of the grammatical encoding process (Schriefers, 1992, 1993; Meyer, 1996; Miozzo & Caramazza, 1999). However, there is uncertainty about *how* much smaller this scope actually is: some theories have argued that only a small amount of advance planning occurs prior to articulation (MacKay, 1987; Jordan, 1990), while more recent theories hold that speakers are able to buffer a larger amount prior to articulation (Wheeldon & Lahiri, 1997; Levelt et al., 1999; Lahiri & Plank, 2010). The amount of buffering depends on a number of variables (such as utterance complexity, working memory load, or time pressure), but allows for the possibility that speakers may be able to encode more than a single Phonological Word at the same time. Nevertheless, these theories stipulate that the minimal unit of encoding in all cases is still prosodically, rather than lexically, shaped. Accordingly, we begin this study with the following research question:

- **What is the post-lexical planning unit in English?**

Attempts to generate experimental evidence for the existence of prosodic constituents, and in particular the Phonological Word constituent, have supported a cognitive process dedicated to the construction of prosodic structures (Schnur et al., 2006; Krivokapic, 2007; Damian & Dumay, 2007; Roelofs, 2008; Schnur, 2011, Wagner et al., 2010; Lahiri & Plank, 2010; Martin et al., 2010; Opperman et al., 2010; Cholin et al., 2011; Konopka, 2012). Although studies in the phonological encoding of Dutch, European Portuguese, and German have elicited strong results for the planning of prosodic units for articulation, very little has been done on English production. Inhoff et al. (1996)

found that naming latencies for English compounds were nearly identical (434 ms) to monomorphemic words (433 ms) in a delayed production task. They attributed this to the morphological composition of the word, and hypothesised that the naming latencies were related to the "initial phase of lexical activation". Beyond this, however, they were unable to explain their results.

Both Levelt (1989, 1999) and Roelofs (1997, 1999) assume a prosodically-shaped unit in their models. Further supportive evidence comes from psycholinguistic tasks employing the prepared speech paradigm, in which subjects complete a question-answer task using the target words in experimental sentences (Sternberg et al., 1978; Wheeldon & Lahiri, 1997, 2002). These studies have elicited results that indicate that speech production latencies are sensitive to a prosodically-shaped unit: the Phonological Word. Recall that a Phonological Word is minimally composed of at least one stressed foot. Clitics (such as auxiliaries, prepositions, pronouns, conjunctions, and articles) are often unable to meet this requirement. They display what Gerlach and Grijzenhout (2000) call 'prosodic deficiency', in that they lack stress and full vowels.

Another linguistic item which portrays this nonisomorphy between syntactic and phonological units is the English compound. In 1977, Liberman & Prince (building on Liberman's 1975 thesis) observed that perceived stress ("relative prominence") in English was depicted one way in lexical compounds and another in phrases. Atkinson-King (1973) reported that the ability to differentiate between compounds and phrases in English is a function of age: that is, younger children were unable to tell the difference between the compound *greenhouse* and the phrase *green house* by stress patterns alone.

Compounds have revealed much about the role of morphology in both comprehension and production; however comparably very little has been done on their phonological status. Jackendoff calls compounds "protolinguistic fossils" that provide clues about the structure of linguistic representation in the mind and grammatical processing (2002: 250). Compounds contain two (or more) morphosyntactic words (e.g. *lighthouse*), which means they also contain two Phonological Words. When treated as individual morphosyntactic units, *light* and *house* are separate phonological units with their own stress assignment (depending on the context in which they occur). When compounded, however, the units merge into a single Phonological Word (*lighthouse*), with main stress falling on the first unit and secondary stress on the second. The question which has rarely been discussed is whether (in terms of post-lexical processing) each of these items are treated similarly: that is, whether the Phonological Word is the unit of post-lexical encoding, or whether it is the lexical unit. The first aim of this study, therefore, is to examine the planning of compound words in native speakers of English.

Our second research question is as follows:

- **What is the post-lexical planning unit for non-native English speakers?**

For many years, the primary focus of non-native speech studies was on language learning and proficiency rather than to the structure and generation of language itself. Dedicated approaches to the cognitive processes of second language production did not appear until the 1960s and 1970s, when researchers like Kolers (1963) and McCormack (1974; 1977) started to look at the structure of semantic stores in bilingual speakers.[2] So what do we know about the production of speech in a non-native[3] language? We know that learning a new language requires the neural systems related to linguistic processing to adapt in order to cope with a new syntax and lexicon (Francis, 1999; Kovelman et al., 2008; Simmonds et al., 2011). We also know that age, method of instruction, and modality (i.e. whether the language is signed or spoken) are all factors in how

[2]Reasons for this are manyfold: Palij & Aaronson (1992) cited a lack of access to adequate pools of participants and a lack of interest in non-native processing as possible explanations. Further issues crop up regarding terminology: the terms "non-native", "bilingual" and "L2" are used without consistency in the literature. For example, in Chakraborty (2011), the term "bilingual" is used to describe speakers who have a wide (and rather confusing) range of fluency: "these two bilingual groups were classified as early exposed and late exposed groups. Both groups had comparable years of exposure to standard American English, ranging from one to 10 years" (492). Finally, there is no standard method of judging proficiency in non-native speakers: researchers regularly use a variety of proficiency scales that are difficult to compare across studies.

[3]In order to avoid the confusion discussed above, I will use the terms "non-native" and "L2" to refer to what Simmonds et al. (2011) call "successive bilingualism": "in which a second language (L2) was learnt after the first (L1) was already established; despite high levels of linguistic proficiency in L2, it is clearly a non-native language because of the persistence of a foreign accent..." (2).

proficient a speaker becomes in their L2.[4]

Speech errors reveal important clues about the structure of the non-native production process: delays, code-switching, incorrect stress patterns, and lexical blends tell us that errors can occur at more than one stage during production. It has been argued that the single most obvious feature of non-native speech is prosody (Hatch, 1983; Dickerson, 1989). In linguistic terms, prosody can be defined as the "suprasegmental" features of a language such as stress, intonation and rhythm (Lehiste, 1970; Fox, 2000). Studies in native language acquisition have shown that prosody is one of the first linguistic structures children are able to use: infants as young as 2 months show sensitivity to prosodic phrases (Hirsh-Pasek et al., 1987; Mandel et al., 1996).[5] Every language has a specific prosodic system that dictates the assignment of features such as stress and syllabification, and errors in these linguistic structures can reveal much about the relative fluency of a non-native speaker. Distinctive features of non-native prosody include deviations in patterns of stress, intonation, timing, phrasing and rhythm (Anderson-Hsieh et al., 1992): for example, as we will see later in this study, native Bengali speakers often import the trochaic Bengali stress pattern onto English words (e.g. [ˈgæzəl] for the final-stressed word [gʌˈzɜl], "gazelle").[6]

Following this, the second part of this study addresses phonological encoding in non-native speakers of English. In particular, we examine the processes that prepare the post lexical metrical structure and Phonological Word frames. We look at the ways in which production is affected by L1 phonological processing, as well as to what extent they are successful in learning the prosodic patterns of the L2 language.

[4]see Snodgrass & Yuditsky, 1996; Brysbaert et al., 2000; Bonin et al., 2001; Davies et al., 2013 and Storkel, 2013, for detailed discussions of these factors.

[5]In particular, they show sensitivity to acoustic cues such as pitch lowering and durational increases (Shukla et al., 2011).

[6]primary stress is marked here in the ˈcvcv format.

1.5 Structure of this Thesis

This thesis is arranged around eight psycholinguistic experiments designed to elicit evidence for the prosodification process in language production. In order to address the crucial underpinnings of both formal linguistic theory and psychological models of language production, two theoretical chapters are included in this study. Chapter 2 presents the framework of prosodic constituent theory, a theory of preordered constituents designed specifically to formalise the relationship between prosody and syntax. In this chapter, we look at the origins of the theory and the studies that form its foundation, focusing in particular on the phonological phenomena associated with complex linguistic items such as compounds and clitics.

Chapter 3 presents a survey and discussion of language production models, in which we will illustrate how the concept of post-lexical encoding grew from a single large-scale process into a multi-faceted procedure involving many different subprocesses. Chapter 4 presents the general methodology of the experiments in this thesis: on account of the same experimental paradigm being used in all eight experiments, it is explained for the sake of brevity here. Chapter 5 contains the results and analysis of Experiments 1 and 2, which investigate compound naming latencies in native English speakers in two different task conditions. In Chapter 6 (Experiments 3 and 4), we investigate how clitics are planned in relation to complex morphosyntactic structures in native speakers. Based on the results of these four experiments, we turn to phonological encoding in non-native English speakers. Chapter 7 presents the results and analysis of the first two tasks of the English study in native Bengali speakers. Chapter 8 follows with two experiments which investigate how clitics are planned in non-native speakers. Finally, Chapter 9 concludes the study by discussing the results of all eight experiments in relation to both linguistic and psychological theories of language production.

2
Prosodic Constituent Theory

2.1 Introduction

Syntax and phonology interact in many ways; as seen in §1.2, these two processes do not always produce representations that are symmetrical in structure. Early generative linguistic theory attempted to account for these asymmetries by proposing separate components for syntactic and phonological information. However more recent theories (e.g. Nespor & Vogel, 1986) have found this approach too simplistic in regards to the phonological component. Instead of consisting of a single process that accepted a syntactic representation as input and generated a corresponding phonological output, Nespor and Vogel saw the phonological component of grammar as consisting of multiple subsystems, each with its own generative function. Here we present a theory of preordered constituents designed specifically to formalise the relationship between prosody and syntax: prosodic constituent theory. Prosodic constituent theory approaches the issue of asymmetry in syntactic and phonological representations by introducing a series of hierarchically-structured units which are each defined by their own rules and principles.

In order to fully understand the current framework of prosodic constituent theory, we must first discuss its origins. Although the beginnings of prosodic constituency framework are often identified with metrical or rhythmic phonological theories of the late twentieth century, its roots were planted much earlier. §2.2 introduces prosodic constituent theory and its current framework, Prosodic Phonology. In §2.2.1, we will look at several contributions to the early understanding of prosodic domains made by eighteenth and nineteenth-century grammarians such as Frans Saran and Henry Sweet. In these contributions, we will see early attempts to analyse speech as a hierarchical grouping of sounds. §2.2.2 introduces the early frameworks of prosodic constituent theory and their relation to the generative tradition, and focuses on the non-linear approaches by Liberman and Prince (1977), Selkirk (1980, 1984), and Nespor and Vogel (1986, 2007). After discussing the foundation of the theory (the Prosodic Hierarchy), we turn in §2.3 to one constituent in particular. Within the frameworks of Selkirk and Nespor and Vogel, there is a proposed domain where morphology, syntax, and phonology interact: The Phonological Word. This is the domain where the asymmetries in syntax and phonology are most visible, particularly in complex morphosyntactic structures containing clitics (§2.3.1) and compounds (§2.3.2). Finally, §2.4 presents a theoretical approach to the treatment of these items based on theoretical work by Selkirk (2011) and experimental work by Wheeldon and Lahiri (1997, 2002).

2.2 Prosodic Constituent Theory

2.2.1 Origins of the Theory

Although the roots of prosodic constituent theory are commonly associated with the metrical grid frameworks of Liberman (1975), Liberman & Prince (1977), and Selkirk (1972, 1980b), the idea that sounds could be organised into hierarchically-arranged groups was first proposed much earlier. In his review of the theory, Booij (1984) makes reference to various hierarchical models presented by the structuralists in the mid-1950s and 1960s such as Hockett (1955), Haugen (1956), and McMahon (1967). But while

Booij's observation is correct in regards to modern theory, Lahiri and Plank (2010) cite considerably-earlier examples of hierarchical approaches in the prescriptive grammars of the eighteenth and nineteenth centuries. Grammarians such as Joseph Steele, Henry Sweet, Eduard Sievers, and Franz Saran all hypothesised that groups of sounds could be organised according to rhythm-related features (such as loudness, intensity, and duration). While their aims were inherently different, these scholars' analyses were strikingly similar to those introduced in the late twentieth century in that they recognised that the spoken organisation of an utterance could be significantly different to its lexical or syntactic structure.

The grammatical tradition of the late eighteenth and nineteenth century was propelled by the prescriptionist movement to standardise the English language (Milroy & Milroy, 1985: 27). Grammars aimed to instruct readers on correct usage; therefore, a large number of these grammars were devoted to the teaching of proper, "correct" syntax (van Ostade, 2008). Out of this movement came manuals and handbooks devoted to proper elocution and delivery. One such manual was Joseph Steele's *Prosodia Rationalis* (1779), which contained a novel system of annotating what he referred to as the "melody and measure of speech".[1] Steele based his system on those used for musical notation[2], scoring features of speech such as loudness, emphasis, and duration on a scale.

He proposed that speech was arranged into "cadences": that is, "emphatic divisions" that reflected the features of spoken language (1775: 35). Figure 2.1 provides an example of Steele's notation system for the phrase, "every sentence in our language":

[1] In fact, "Prosodia Rationalis" was the short title for "An Essay Towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols".

[2] The association with the two structures is not unusual, considering the similarities between speech and music. The metrical structure of music even has a familiar shape: it consists of alternations of strong and weak beats, hierarchically-organised according to accent levels (Palmer & Kelly, 1992; Jackendoff, 2009). Cf. Hayes (1989):

> "As in music, phrasing in language is hierarchical: the lowest units are grouped into small phrases, which in turn are grouped into larger phrases, and so on through levels." (201)

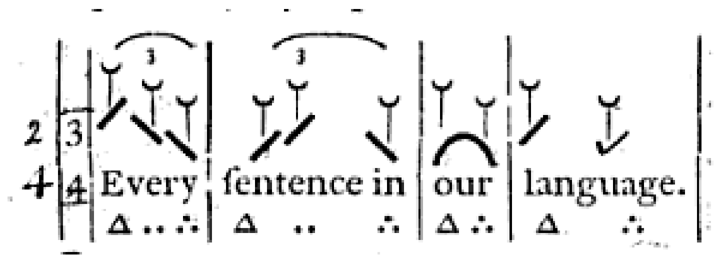


Figure 2.1: Steele's Notation System (1775, 29)

Each cadence (whose boundaries are marked here by vertical bars) contained marks for emphasis, syllable weight, and tone. The triangle notation, located at the left edge of each cadence, denoted the part of the structure that carried the most prominence (or intensity). Unlike other elocutionists of the eighteenth century whose aims were to prescribe ideal or correct speech patterns, Steele focused almost entirely on the structures of spoken language as they existed in natural (connected) speech (Newman, 1951). As such, we see Steele attempting to reconcile the fact that certain features of spoken language are at odds with its orthographic structure. Note the cadence containing "sentence in": Steele writes,

> In this alternation, the monosyllable *in* (one of our pliant syllables) which before was *heavy* and *acute*, becomes *light* and *grave*... (1775: 29)

This observation is a significant one for our analysis: Steele realised that the monosyllabic function word "in" became weak ("pliant") in spoken language, and therefore attached leftwards to the word "sentence". In this notation, Steele is organising speech not in orthographic (and therefore lexical) patterns, but in phonological patterns.[3]

The observation of the asymmetry of spoken and written language was made repeatedly in these early analyses: in his *Elementarbuch des gesprochenen Englisch* (1886) (a grammar of English for German learners), Henry Sweet emphasised the importance of spoken English over its orthographic representation and often deferred to aspects of language related to what he called *raschem sprechen* ("rapid speech") (p. lxv). With his

[3]This was, of course, not the first prosodic notation system to be proposed. The study of prosody has its philological roots in the rich tradition of grammatical treatises: Aristophanes of Byzantium (ca. 200 B.C.) is credited with the invention of prosodic notation (Probert, 2006).

grammar of English (aimed at German learners), he included exercises presented in an intriguing format:

(2.1) *-pijpl juwsttəθinkði əəðwəzə kaindəv flæt keik*

'People used to think the earth was a kind of flat cake.' (1885, p. lxv)

Rather than being arranged in a regular orthographic style, the sentence in example 2.1 is presented in what Sweet perceived to be the sentence's "phonetic form", where words were grouped together to achieve a more "normal" representation of regular speech (Sweet, 1885: iii).

Sievers (1901) also favoured the organisation of sounds by spoken-language features in his *Grundzuge der Phonetik*: along with other word and phrasal classes, he proposed that utterances were organised by *Satztakte* ("stress-groups)". Furthermore, within each of these groups, the individual syllables of a sentence did not have the same stress value; instead they were subject to a rule that combined weak syllables with strong syllables.

Franz Saran (1907) examined rhythm and meter in spoken language in his *Deutsche Verslehre*; in particular, he focused on the ways in which sounds combined in certain constructions, but diverged in others. In relation to this, he directed his readers to Sweet's observations about the differences between the phonetic and etymological forms of a sentence:

> *Es ist eine höchst bedeutsame Entdeckung von Sweet, daß die phonetische Gliederung eines gesprochenen Satzes mit der logisch-etymologischen Zerlegbarkeit desselben in Wörter nicht verwechselt werden darf....wann werden die Wörter und die syntaktischen Gruppen so zerrissen?*[4]

In investigating where this division between syntactic groups and words occurred, Saran proposed a hierarchically-organised system of sounds, categorised by strength, loudness, and syllabification. These groups had descriptors such as *die Silbe* ("the syllable"), *das Unterglied* ("the submember"), *das Glied* ("the member"), *die Reihe* ("the row"), *die*

[4]"Highly significant here is Sweet's determination that the phonetic division of a spoken sentence must not be confused with its logical-etymological dissection into words.. . . when are the words and syntactic groups broken up in this way?" (Saran, 1907: 80)

Kette ("the chain"), *das Gebinde* ("the binding"), and *das Gesätz* ("the utterance"). These groups were arranged according to their behaviour in regards to sound rather than syntax: the smaller groups (i.e. *die Silbe*) referred to structures smaller than a word, and the larger above it (*das Gesätz*). The middle structures (i.e. *die Reihe* and *die Kette*) dealt especially with features of connected speech: that is, where Saran's words and syntactic groups did not always meet. As we will see in later sections, this system has a great deal in common with modern constituent-based theories: in fact, it is here that we have our first rudimentary prosodic hierarchy.

It is important to acknowledge these early treatments of sound groups because they underlined the non-isomorphy of syntactic and phonological structure. So impressed were the early grammarians by their findings that they defaulted to theories that thoroughly rejected **any** interaction between the syntax and phonology (Lahiri and Plank, 2010).[5] However, as we will discuss in the following section, evidence arises that requires at least some relationship between phonology and syntax.

2.2.2 Frameworks

This section will introduce the major incarnations of prosodic constituent theory, briefly discussing each framework's contribution to our current understanding of the theory. Beginning with the seminal work of SPE, we will see how issues related to the non-isomorphy of syntax and phonology have enriched different non-linear theories.

2.2.2.1 Chomsky and Halle, 1968

Now that we have established the origins of the theory, as well as how the asymmetry of sound and syntactic structures guided the early frameworks, we can turn to prosodic constituent theory itself. In their pioneering study of modern linguistic theory,

[5]cf. Saran: *Wie die 8 Schwerestufen den Silben selbständig gegenüberstehen, weder aus den lexikalischen noch syntaktischen Eigenschaften der Worte abgeleitet werden können, so auch das System der akzentuellen Gruppen*. ("Like the 8 degrees of stress, the system of accentual groups is also independent of syllabic considerations, impossible to derive from the lexical or syntactic characteristics of words") (1905: 95).

Chomsky and Halle (1968) proposed that lexical representations contained separate components for syntactic and phonological information. Once the syntactic components prepared the syntax of the utterance, the surface structure was fed into the phonological component, where a number of "readjustment rules" prepared the output phonological representation (1968: 9).

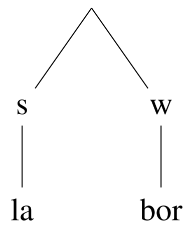
Like the early grammarians, Chomsky and Halle recognised that syntactic and phonological structures did not always coincide: they viewed this as a result of a disparity in the input/output representations between the two components. But unlike the early grammarians, they argued that there had to be some (in fact, a "very significant") interaction between syntax and phonology (1968: 9). One example of this interaction involved the assignment of word stress: Chomsky and Halle theorised that the readjustment rules associated with the phonological component erased the original, syntactic bracketing in the surface structure, and issued new bracketing that generated the correct stress pattern. It is in this operation of stress assignment that the roots of modern prosodic constituent theory are located.

2.2.2.2 Liberman and Prince, 1977

Liberman & Prince (1977) followed Chomsky & Halle in viewing the syntactic and phonological components of language as separate processes, but proposed that English stress was not simply assigned to a single syllable or segment as in *the Sound Patterns of English* (henceforth referred to as SPE). As evidence for this argument, they referred to structures that contained the same lexical units but displayed different stress patterns (e.g. the compound *blackbird* and the noun phrase *black bird*). They theorised that segments had to be marked for stress at more than a single level of phonological representation (1977: 270). For example, the assignment of stress in the disyllabic word "labor"[6] is fairly straightforward in example 2.2 below (1977: 22):

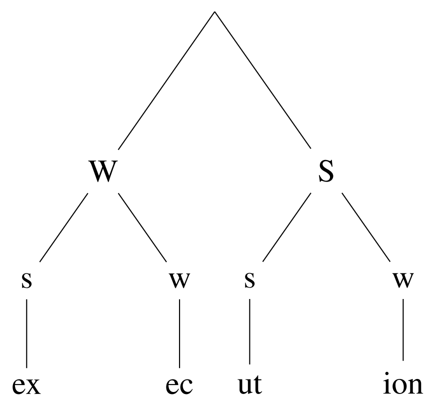
[6] Liberman and Prince's American English (AmE) spelling is retained here.

(2.2) "labor"



Stress is simply assigned to two nodes in this parentless tree, with the first syllable marked for heavy stress and the second for weak. However, Liberman and Prince found that a single level was insufficient on which to illustrate the assignment of stress in multisyllabic words (such as "execution" in example 2.3 below):

(2.3) "execution"



Instead, Liberman and Prince assigned stress on higher-level nodes (W,S) in addition to the individual syllable nodes (s,w,s,w). They called these higher nodes the "metrical foot nodes", where the "relative prominence" was marked for the word as a whole. The parent node of the entire structure remained empty in this early analysis. In the metrical foot nodes, the right node could only be marked strong if it dominated a branching node: this constituted Liberman & Prince's "word rule":

(2.4) **Word Rule (1977: 268)**

In a pair of sister nodes [N_1N_2], N_2 is *s* if and only if it branches.

That is, the second morpheme of a structure could be strong only if it contained a branching structure. Like the early grammarians, Liberman and Prince found that hierarchically-organised groups dealt well with the issues related to the non-isomorphy of the syntactic and phonological components of grammar. They recognised the necessity for different levels and constituents to handle different phonological features (such as syllable weight and stress). Furthermore, as Hayes (1980) noted, they were able to explain why stress rules often "behave differently" to other rules (11). The realisation that analyses of word stress in English require multiple layers in order to successfully depict stress assignment was a critical underpinning of prosodic hierarchical theory.

2.2.2.3 Selkirk 1980, 1984

Liberman and Prince's work formed the foundation for a number of major non-linear approaches to stress assignment, e.g. Halle and Vergnaud (1978), McCarthy (1979), Selkirk (1980, 1984, 1986, 1995), and Hayes (1980, 1985). These approaches focused heavily on the roles played by the syllable and foot in word stress, and in particular on the internal structures of these units. Selkirk's (1980a,1984) theory of English word stress shared many features of Liberman and Prince's analysis, such as the binary-branching tree structures and hierarchically-arranged nodes.

But while Liberman and Prince's analysis assigned stress to binary-branching trees in terms of the segmental feature [± stress], Selkirk's analysis followed approaches which argued against the inclusion of this feature in phonological theory (cf. Halle and Vergnaud, 1978 and McCarthy, 1979). Instead of using the segmental feature to depict stress in phonological structures, these analyses relied upon the properties of the prosodic categories themselves (Selkirk, 1980: 574). However Selkirk found these analyses lacking, in that they deferred to the same types of binary-branching tree structures as Liberman and Prince. Instead, Selkirk incorporated Liberman and Prince's prosodic category labels into six hierarchically-organised structures called "constituents", making up the Prosodic Hierarchy (Figure 2.2):

Phonological Utterance (UP)

Intonational Phrase (IP)

Phonological Phrase (φ)

Prosodic Word (ω)

Foot (Σ)

Syllable (σ)

Figure 2.2: Selkirk's (1980) Prosodic Hierarchy

Each constituent in the Prosodic Hierarchy was composed of one or more units of the immediately lower category, e.g. the Phonological Utterance was made up of one or more Intonational Phrases, which was made up of one or more Phonological Phrases. At the word level, Selkirk introduced three constituents: the Syllable, the Foot, and the Prosodic Word: these constituents, she argued, constitute the fundamental building blocks of a phonological representation. In this sense, Liberman and Prince's levels were expanded into individual, self-contained groups. This system, the Prosodic Hierarchy, is referred to as the foundation of the current framework of prosodic constituent theory: Prosodic Phonology.

One of the major theoretical tenets that underlie prosodic constituent theory is the rejection of boundary-based analyses in favour of hierarchically-arranged ones (cf. Revithiadou, 2011). In Selkirk's frameworks, each prosodic constituent was subject to specific rules associated with its domain: the application of said rules was dictated by the Strict Layer Hypothesis (abbreviated herein as SLH):

(2.5) The Strict Layer Hypothesis (Selkirk, 1984):

> A constituent of category-level *n* in the prosodic hierarchy immediately dominates only (a sequence of) constituents at category-level *n-1* in the hierarchy.

The SLH mandated that a constituent could only dominate the constituent beneath it: this meant, for example, that a Phonological Phrase could not skip a level and dominate a Foot. In this way, the SLH universally restricted the size and behaviour of the constituents, and in doing so, excluded the need for boundary-related phonological rules (cf. Hayes, 1989).

2.2.2.4 Hayes, 1984, 1989

Hayes' (1984, 1989) Prosodic Hierarchy was also shaped by the constraints of the SLH; however, he argued for only five constituent levels: the Utterance, the Intonational Phrase, the Phonological Phrase, the Clitic Group, and the Word. The Clitic Group aimed to encompass any content words plus any "contiguous grammatical words in the same syntactic constituent" (1989: 207). A Clitic Group would contain a "host" (the content word) and any number of clitic words to the left or right of the host. Ergo, the sentence "He kept it in a large jar", would contain 3 Clitic Groups:

(2.6) He kept it in a large jar.

[He kept it]$_C$ [in a large]$_C$ [jar]$_C$.

Hayes' analysis, while primarily concerned with prosodic theory as applied to poetic meter, is notable because of its attention to the little pieces (i.e. clitic words). The issues related to these items continue to raise questions for the Prosodic Hierarchy and its constituents throughout different analyses, as we will see in Nespor and Vogel's framework in the following section.

2.2.2.5 Nespor and Vogel, 1997, 2007

In line with previous constituent-based analyses, Nespor and Vogel also found Chomsky and Halle's initial analysis lacking in the depth they felt was necessary to account for the multifarious phonological features of language. In particular, they disagreed with the proposal that any information not included in the surface syntactic structure could therefore not affect the application of phonological rules (1986: 4). Such a claim, they argued, was weakened by rules that drew on information stored elsewhere in the utterance, e.g. the semantic information encoded in the rule for "Linking-r ... in the RP style of British English" (1986: 4). In this rule, the [r] sound in word-final position spreads across two semantically-related sentences (as in their example: "There's my mothe[r]. I've got to go."). The fact that this occurs without any syntactic relation requires an analysis that allows for an interaction not only between phonology and syntax, but also between phonology and semantics.

Nespor and Vogel (1986) saw the phonological component of language as composed of interacting "subsystems" that drew on different aspects of metrical grid theory, Lexical Phonology, Autosegmental Phonology, and Prosodic Phonology. Their theory of Prosodic Phonology existed as a subsystem of the phonological component in grammar, and it interacted in numerous ways with other subsystems through the application of phonological rules. The phonological rules that dictated the phonological patterns were the same as in Selkirk (1980); they also replaced the stress assignment rules seen in Liberman and Prince with what they called a "simpler convention":

(2.7) Principle 4 of Nespor and Vogel's Mapping Rules (1986: 7):

> The relative prominent relation defined for sister nodes is such that one node is assigned the value strong (s) and all the other nodes are assigned the value weak (w).

Other notable features of the theory included assigning all morpho-phonological processes to Lexical Phonology (e.g. Kiparsky, 1982), proposing *n*-ary branching foot trees (instead of binary, cf. Hayes, 1980), and the use of multiple types of source material.[7]

Nespor and Vogel proposed a seven-level Prosodic Hierarchy, identical to Selkirk save for the addition of the Clitic Group (Figure 2.3 below).

[7]Nespor and Vogel collected copious amounts of standard Italian, Modern Greek, and English speech data as part of the empirical survey of languages in their study. Other data came from literary sources and native speaker judgments in Spanish, Dutch, French, Japanese, Hungarian and many other languages.

Phonological Utterance
|
Intonational Phrase
|
Phonological Phrase
|
Clitic Group
|
Phonological Word
|
Foot
|
Syllable

Figure 2.3: Nespor and Vogel's (1986) Prosodic Hierarchy

Following Hayes (1984, 1989), the Clitic Group was introduced by Nespor and Vogel as a level between the Phonological Word and Phonological Phrase that handled all "clitic-related phenomena", such as affixes, clitics, and other function words. Multiple issues (which will be elaborated upon in the following section) have arisen in relation to this constituent, and subsequently it has largely fallen out of favour (cf. Inkelas, 1991). In 2007, Nespor & Vogel proposed a new constituent: the Composite Group. Located at the same level as the former Clitic Group, this constituent was expected to handle clitics as well as an additional structure that presented complications for the Phonological Word domain: compounds. The Composite Group aimed to handle features such as stress, attachment, and external prosodic framing; however, as we will show in the following section, there are problems with this group as well.

As is clear from the analyses by Hayes (1989) and Nespor and Vogel (1986, 2007), some the issues have arisen with the regards to the Prosodic Hierarchy; a number of these are situated in or around the Phonological Word constituent.[8] In particular, clitics and compounds appear to cause the majority of problems and so it is with these structures that we will continue our discussion below.

2.2.3 Interim Summary

In this chapter, thus far, we have seen the evolution of a theory of preordered constituents from the prescriptive grammar-based tradition of organising words by sound features, to a detailed hierarchical framework that formalises the relationship of the morphosyntactic and phonological components of language. The asymmetry of spoken and written language has underlined prosodic analyses for centuries now, and attempts to assign prominence and stress have led to the proposal of a theory of preordered, hierarchically-organised constituents. Prosodic Phonology proposes a multi-dimensional approach in which language is treated as chunks of hierarchically-organised domains characterised by specific phonological and phonetic features. It is a theory that is content in working with other theoretical subsystems, such as Autosegmental Phonology and Lexical Phonology. As mentioned above, one constituent in particular is particularly relevant for this study: the Phonological Word. The remainder of this chapter will be devoted to a discussion of this constituent's features and functions, as well as two structures that raise issues for it in recent analyses: clitics and compounds. We argue, following recent experimental and empirical evidence, that both of these complex structures involve an additional, recursive constituent layer.

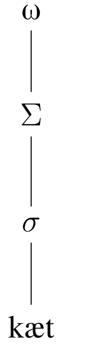
[8]While the structures of the outer constituents of the Prosodic Hierarchy (e.g. the Utterance and the Syllable) are generally undisputed, the middle constituents in general have not fair as well. Some theorists (Beckman & Pierrehumbert, 1986 and Silverman et al., 1992) have argued against the inclusion of certain constituents, while others (Condoravdi, 1990; Bresnan & Kanerva, 1989; Kanerva, 1990) have introduced new constituents (e.g. the Minimal Phrase, and the Focal Phrase) into the Prosodic Hierarchy.

2.3 The Phonological Word

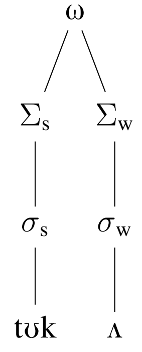
The Phonological Word is the constituent that dominates the foot, and in turn is dominated by the larger phrasal categories in the prosodic hierarchy (i.e. the Clitic/Composite Group in Nespor and Vogel, and the Phonological Phrase constituent in Selkirk). In Selkirk's framework, this constituent was labeled the "Prosodic Word", and it was made up of one or more stress feet (Σ).[9] This unit, as well as the Stress Foot and Syllable, played crucial roles for the assignment of word stress in English. These lower-level constituents, she theorised, all inhabited the same syntactic domain in English: a simple, non-branching stem. Crucially, this stem corresponded to a morphological structure that was smaller than a syntactic word (1980: 570).

In Nespor and Vogel's framework of Prosodic Phonology, the Phonological Word also dominated the Foot constituent and each Foot was exhaustively dominated by a single Phonological Word (1986: 109). The mapping rules, referenced in §2.2.2, reorganise morphological structure into non-morphological constituents: in doing so, they constitute the nonisomorphism of the phonological and syntactic components of the utterance (1986: 110). A Phonological Word contains, at the very least, a stressed foot (example 2.8 below: "cat"). At its largest, it is a single lexical word combined with unstressed function words (such as in example 2.9 below, "took a").

(2.8) **"cat"**



(2.9) **"took a"**



[9]"a sequence of one or more Σ' joined in a right-branching structure (= the principles of construction, C_{ω})"(1980: 565).

When spoken in normal (or rushed) speech, the structure in example 2.9 ("took a") becomes a single Phonological Word, with "a" reducing to a [ʌ] sound and attaching (**cliticising**) to the content word "took": [tʊkʌ]. Cliticisation demonstrates exactly how the number of lexical words can differ to the number of Phonological Words in an utterance. This behaviour, while a predictable feature of spoken speech, raises certain issues for the Phonological Word domain.

2.3.1 Clitics and the Phonological Word

Cliticisation is ordinarily seen with function words, such as auxiliaries, prepositions, and pronouns. By reducing and attaching to neighbouring words, function words can form entirely new structures such as "could've" in the English contraction for "could have" (example 2.10 below).

(2.10) a. could have > could've ([kʊd hɑv] > [kʊdəv])

The auxiliary "have" collapses in to "could" in regular or rushed speech, resulting in the loss of the initial [h] and reduction of the vowel sound to [ə]. This is a common operation in spoken English, and almost always involves leftward attachment of the auxiliary to a host word:

(2.11) a. I am > I'm

b. what are > what're

c. those will > those'll

However, not all clitics behave like those discussed above. In his 1977 monograph *On Clitics*, Arnold Zwicky drew attention the relationship of syntax and phonology in cliticised structures, which caused him to propose three different classes of clitics: simple, special, and bound.[10] Zwicky's classes were categorised based on how the clitics behaved structurally; e.g., how they were ordered, what rules applied to them, and how they acted with other morphemes. Auxiliaries, personal pronouns, determiners, "dummy nouns", prepositions, conjunctions, and adverbial words were regularly classed as "simple clitics": these consisted of unstressed, phonologically-deficient words that attached or relied upon a host word (such as "have" in example 2.10 above). Special clitics were those that acted as variants for stressed, free forms, such as those regularly found in French:

(2.12) a. Je *le* vois. ("I see him")

b. Il *te* donne ("he/it gives you")

The third category contained bound morphemes, such as those often found in Latin:

(2.13) a. arma virum*que* ("of arms and the man")

b. *vobis*cum ("with you", pl.)

These clitics, Zwicky wrote, showed "considerable syntactic freedom", in that they could enhance an entire noun phrase despite attaching only to one unit of the phrase (cf. example (a) in 2.13 above, where the *que* ("and") morpheme applies to both *arma* ("arms") and *virum* ("the man")). Central to Zwicky's argument for different clitic types was the claim that rules apply differently to the three clitic types; that is, some clitics exhibit a very close relationship with their syntax, while others deviate quite significantly from the syntax. In Zwicky's classification system, special clitics exhibited "special syntax and opaque phonology" while simple clitics exhibited "ordinary syntax and ordinary phonology" (1977: 4).

In looking at the different ways in which these clitics behave, we are reminded of Saran's question (cf. §2.2.1): where do the syntax and phonology of the utterance diverge?

[10] Although we can certainly agree that the modern study of clitics was launched by Zwicky, these special items have received linguists' attention from a much earlier date: for example, Sweet noted the special behaviour of "subordinate words" (*untergeordnete Wörter*) such as "have" in his 1891 work.

That is, what causes some clitics to deviate from the structure present in the syntax, while others stay true to it, as in example 2.14 below (taken from Kaisse, 1983):

(2.14) a. Let's go where they (*are/ *re*)!

b. He (*is/ *s*), I should think, quite tired after his journey.

We know that certain syntactic environments in English block auxiliary reduction (cf. Labov, 1969; Lakoff, 1970): in the first example in 2.14, reduction of *are* is blocked by the deletion site for the second instance of *going*. Reduction also fails in situations where an element has been added after the auxiliary, as in the second example in 2.14. Lahiri and Plank (2010) also offer some clues to the origin of this phenomenon: they direct us to historical changes that occurred in Germanic languages, such as the reduction of the reflexive pronoun *sik* ("himself") in Old Norse. In some cases, reflexive Old Norse pronouns became reduced (e.g. *sik* > -*sk*) and attached to the preceding verb, as in example 2.15 below:

(2.15) *festrin* **vannsk** *eigi til jarðar*
rope touch-REFL not to earth
"the rope was not long enough **to touch** the ground" (Fms. ix 3 (5 37)).

This was not an automatic process, however: for example, *sik falla* is not reduced to **fallask* in this line from the skaldic poem *Þorsdrapa* below.

(2.16) *Harðvaxnar let herðir, hall-lands of* **sik** **falla**
mightily-swollen let hardens, stone-lands over himself fall
"the hardened one let the waves **wash over him**" (Þorsdrapa (1851), 7:1-2)[11]

[11] Skaldik poems such as *Þorsdrapa* are characterised by very strict syntactic rules and extensive use of "kennings", highly-structured figurative devices that usually consist of a base word and a qualifier (e.g. *harðvaxnar"*- lit. "the mightily-swollen ones" for "waves" in example 2.16 above) (McTurk, 2008: 486).

So why does *sik* reduce in example 2.15 but retain its strong form in example 2.16? One hypothesis would be that the verb *falla* requires strong forms of pronouns, and therefore there is no reduced form. But this is easily disproven, as multiple examples of the word *fallask* are easily found in search of the Old Norse text corpus (cf. examples 2.17 and 2.18 below):

(2.17) *Þá lét Loki* **fallask** *í kné Skaþa*
then let Loki fall-REFL on knees Skaþi
"Then Loki let himself drop onto Skaþi's knees... (*Skáldskaparmál*, Snorri Sturluson 1875, p. 2)"

(2.18) *oft sitjanda sögur um* **fallask** *ok liggjandi lygi um bellir.'*
Often sitting-one tales around fall-REFL and lying-one lies around prone
"Often the yarns of the seated one break down, / and one who is lying down is prone to tell lies" (Þrymskviða, v. 10)[12]

The answer lies in the syntactic structure: in example 2.16, it seems that cliticisation[13] is blocked in one direction by something in the sentence's syntax. The syntactic structure of the verb phrase in example 2.16 (*sik falla*) is [REFL VERB], while in examples 2.15 and 2.17, the structure is [VERB REFL]. Unlike modern Germanic, the default syntactic structure of Old Norse (and other early Germanic languages) was trochaic ([GRAM_1 LEX_1][GRAM_2 LEX_2]) (Lahiri and Plank, 2010). Despite this, there are many examples of encliticisation and suffixation in the language; therefore, while the syntax preferred [GRAM LEX], it seemed perfectly content to produce constructions with both (GRAM LEX) (e.g. *falla sik*) and (LEX GRAM) (e.g. *sik falla*) (see example 2.19 below).

[12]Translation of text from Jónsson (1954) by Dr. Anthony Adams, personal communication, 9 December 2015.

[13]The use of the term "cliticisation" in regards to reflexive elements (aka. Middle Voice) in Old Norse is contentious: Kemmer (1993) argues that *-sk* had already reached affix status in Old Norse, as opposed to an unstressed, cliticised version of the pronoun (186). However, Faarlund (2004) disagrees, arguing that the reflexive pronoun forms qualify perfectly well as Old Norse clitics, in that they alternate with independent words and they act as independent elements in the syntax (functioning as an argument). Ottosson (2008) assigns the reflexive form of *sik* to Zwicky's "special clitic" class, arguing that it displays a syntactic distribution different to its corresponding full form (189).

(2.19) Old Norse *falla* + reflexive pronoun full form:

GRAM LEX → (GRAM LEX) (LEX GRAM)

falla + sik → (sik falla) (falla sik)

But the process of cliticisation was blocked in the syntactic structure [GRAM LEX]; *sik falla* never reduces to **skfalla*:

(2.20) Old Norse *falla* + reflexive reduced form:

GRAM LEX → *(gram(LEX) (LEX)gram)

falla + sik → (*sk(falla) (falla)sk)

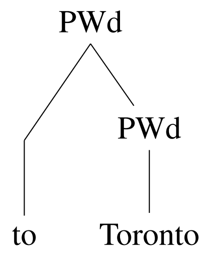
Therefore, the default syntactic structure, while perfectly conducive to encliticisations, did not allow the reflexive pronoun *sik* to reduce to *-sk*. Here we have evidence of the phonology working in opposition to the syntax of the utterance. This "asymmetry" of clitic attachment is prevalent throughout Germanic, and it further underlines the relationship between phonological and syntactic structure.

Crucially for this discussion, boundary-based theories such as SPE assumed that clitics could not attach to boundaries in the same way as they would a prosodic constituent. However, Selkirk theorised that function (grammatical) words such as those discussed above possess a fundamentally- different phonological structure than lexical words: specifically, they are invisible to the mapping processes that associate syntactic structure with phonological structure (1984: 226). That is, when the syntax of an utterance was translated to phonological structure, the metrical grid's alignment operations would simply ignore any function words in the utterance.[14] In 1995, Selkirk proposed that function words functioned either as individual Phonological Word units, or prosodically-deficient function words called a "prosodic clitics". These function words regularly exhibit specific

[14] Although her 1984 publication takes a detour from the Prosodic Hierarchy in order to focus on the metrical grid, Selkirk's work still represents a non-linear (autosegmental) approach which was in tune with other post-SPE approaches that sought to move away from the linearly-oriented concept of boundaries.

phonological features, such as the post-lexical aspiration of initial voiceless obstruents in constructions such as "to Toronto" below:

(2.21) tʊ tərɒntoʊ > tətərɒnoʊ



Selkirk further divided the category of "prosodic clitic" into three different types, as characterised by their prosodic structure:

(2.22) **Clitic Types (Selkirk, 1995)**

prosodic word: $((\text{fnc})_{\text{Pwd}}(\text{lex})_{\text{Pwd}})_{\text{PPh}}$

free clitics: $(\text{fnc }(\text{lex})_{\text{Pwd}})_{\text{PPh}}$

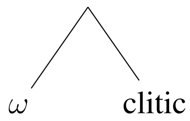
internal clitics: $((\text{fnc lex})_{\text{Pwd}})_{\text{PPh}}$

affixal clitics: $((\text{fnc}(\text{lex})_{\text{Pwd}})_{\text{Pwd}})_{\text{PPh}}$

In these cliticised structures, the nonisomorphy of syntax and phonology is particularly noticeable: while the syntactic structure is simply "[Fnc Lex] XP", the possible options for clitic type listed above illustrate the complexity of the phonological structure. The structure in example 2.22 represents the "free clitic" type.

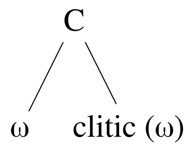
Recall that, in their 1987 framework, Nespor and Vogel (and Hayes, 1989; Vogel, 2009) assigned clitics to the Clitic Group. The domain of the Clitic Group constituted a Prosodic Word plus any adjacent Prosodic Words containing either a directional clitic (enclitic or proclitic) or a clitic which could not be shared amongst categories (1986: 162). They argued that the specific behaviour (such as that discussed above) of clitics required them to be placed in a separate constituent. As we have seen, clitics often exhibit specific behaviour when attaching to a host, such as direction of attachment (cf. examples 2.15 and 2.16). Nespor and Vogel argue that this behaviour generates a representation (cf example 2.23) that is fundamentally-different to that of the Phonological Word constituent.

(2.23) host words and clitics



But this presented problems for the Prosodic Hierarchy: recall that a Phonological Word must have at the very least a stressed foot made up of a strong syllable that contains two morae. Clitics often lose stress when attaching to the host word; therefore they do not meet the requirements of Phonological Words. If we return to the Prosodic Hierarchy and its constraints, we are reminded by the SLH that nodes cannot dominate other nodes of the same type. Therefore the structure in example 2.24 below is in violation:

(2.24) clitic group dominating an encliticising structure



On account of this (and other issues), the Clitic Group was largely rejected. However, concerns related to clitics and their position in the Prosodic Phonology framework did not diminish; furthermore, the position of other complex structures began to be questioned, namely compounds.

2.3.2 Compounds

As linguistic items, compounds have received a great deal of attention due to their unique structure, particularly in relation to noun phrases (cf. Chomsky & Halle, 1968; Ladd,1984; Cinque, 1993; Fabb, 1998; Badecker, 2001; Booij, 2005; Fiorentino & Poeppel, 2007). In 1989, Hayes noted that compounds presented a problem for the

Prosodic Hierarchy: "Obviously many questions remain. For example, there is little evidence to determine what phrasing should be assigned to compound words" (220).

English compounds exist in a variety of word classes, and are often identified by special semantic, morphological, and phonological criteria that makes them different from phrases; however as we will see below, the distinction between the two structures is not always transparent. The categories into which the different types of compound words have been assigned are numerous and complex; the issue of what actually constitutes a compound is not an easy one. Existing categories are generally based on three criteria: headedness, inseparability, and stress (cf. Lieber, 2006).

The criterion of headedness refers to the location of the head, or root, of the compound. In English, endocentric compounds tend to exhibit heads in the rightmost position, where they modify the rest of the compound:

(2.25) a. light*house*
b. sun*flower*
c. hand*bag*

However, not all compounds contain obvious heads: exocentric ("bahuvrihi") compounds exist quite happily without any "formal" heads (cf. Benczes, 2004). That is, although the compounds contain rightmost heads, their meaning cannot be easily gleaned:

(2.26) a. hammerhead
b. jailbird
c. scarecrow

Another commonly-referenced criterion of English compounds is their inseparability; elements cannot be inserted between the two compound constituents. For example, while it is possible to insert another element into a phrase:

(2.27) a. light *cotton* coat

b. old *red* wagon

c. happy *little* tree

one cannot do the same with a compound word:

(2.28) a. *light*red*house (for *a red lighthouse*)

b. *ground*fuzzy*hog (for *a fuzzy groundhog*)

c. *play*safe*ground (for *a safe playground*)

Furthermore, the first constituent of an English compound generally rejects modification such as plural marking:[15]

(2.29) a. *hand*s*bag

b. *foot*s*print

c. *black*s*board

Lastly, word stress is used as a phonological criterion for distinguishing compounds from phrases in English (Chomsky & Halle,1968; Liberman & Prince, 1977; Halle & Vergnaud, 1987). [16] Liberman and Prince (1977) noted that larger "compound"[17] structures such as "labor union"[18] posed problems for their word stress rule (see example 2.25):

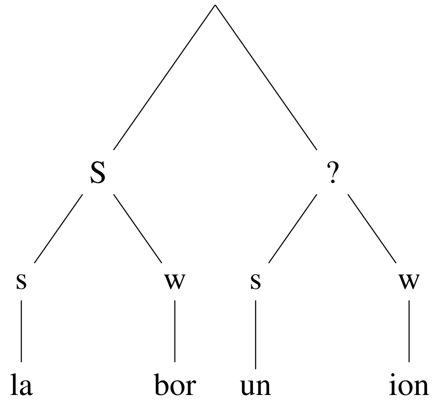
[15]This characteristic is clearly language-dependent of English: as we will see later in this study, Bengali compounds can exhibit case marking on the first constituent, e.g. *gayehalud*, 'smearing tumeric paste on body' (Dash, 2015: 216).

[16]But there are exceptions to the regular stress patterns of compounds and phrases: cf. Fudge (1984), Liberman and Sproat (1992), Sproat (1992), Giegerich (2004), Plag (2006); Plag et al. (2007), 2008. For one, native English speakers often vary stress depending on dialect: there are examples of compounds to which compound stress doesn't seem to apply (e.g. *Madison Ávenue*) (cf. Plag, 2006). Furthermore, compounds that contain a present or past participle as the second stem are often stressed on the right constituent: *man-MADE* (Marchand, 1969: 15).

[17]It could be argued that "labor union" is a phrase and not a compound, however the analysis also works for multi-syllabic compounds such as "riverbed" or "teeter-totter".

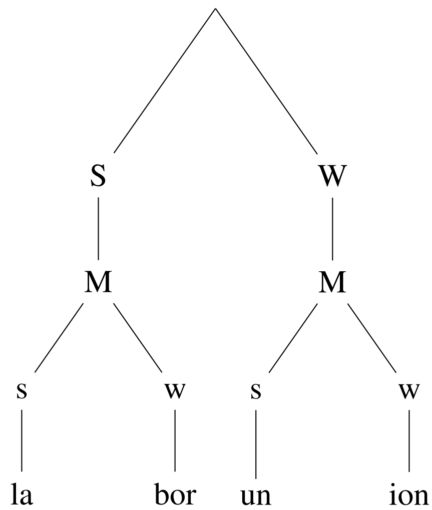
[18]again their AmE spelling is retained here.

(2.25) "labor union"



Following the word rule, both morphemes (*labor* and *union*) should be strong. However, we know this is not the case: in regular speech, stress falls on the leftmost syllable, then is assigned from left to right. Here, Liberman and Prince proposed a special level, " *mot*", which they defined as the "prosodic level" (1978: 270). Now the compound could be marked with correct phrasal stress without breaking the word (or compound) rule (example 2.26):

(2.26) "labor union"



In 1987, Nespor and Vogel assigned compounds to the Phonological Word constituent, claiming that a single compound constituted a single Phonological Word. Once again, the definition of a Phonological Word raised problems for this analysis: compounds contain two (or more) morphosyntactic words, which (following the definition of the

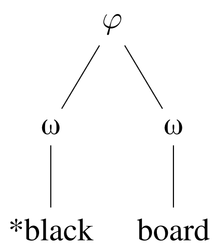
Phonological Word) means they also contain two Phonological Words. In isolation, each of these words acts as a separate phonological unit with its own stress assignment (depending on the context in which it occurs). When compounded, the units form a larger structure, with the main stress falling on the first unit and secondary stress on the second (shown in a metrical grid in example 2.27):

(2.27) Stress of a Compound:

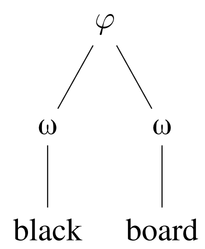
```
  x
  x    x
blackboard
```

The second syllable of the compound still carries stress and cannot be considered prosodically-deficient: unlike clitics who reduce and attach to lexical hosts, these morphemes are well-formed Phonological Words in their own right. Nor can compounds be treated as Phonological Phrases; doing so would leave nowhere to designate the compound stress representation:

(2.28) *compound: BLACKboard



(2.29) phrase: black BOARD



The structures in examples 2.28 and 2.29 are indistinguishable, and the different stress contours are unidentifiable.

In their 2007 revision of the framework, Nespor and Vogel assigned compounds to the Composite Group. Again, they argued that a new constituent was necessary on the basis that certain phonological rules only applied to the outer structures of these words. Furthermore, they maintained that the theory of Prosodic Phonology is built on the concept that prosodic constituents are hierarchically-organised and therefore not capable of recursivity, and a violation of this contradicts the basic tenets of the ideology itself.

A specific argument in favour of the Composite Group involves a mismatch in the phonological processes occurring at the word and compound level. Vogel (1994) refers us to the Italian compound *girasole* ("sunflower", lit. "turning towards the sun"). Italian exhibits regular voicing of intervocalic [s] within prosodic words, e.g. *izola* for *isola*, "island":

(2.30) **Intervocalic Voicing Rule**: s -> z / V____V

However this is blocked in compounds such as *girasole* (**girazole*) and *sputasangue* (**sputazangue*) ("bloody nose"). Vogel argues this blocking suggests that the outer structure of the compound is fundamentally different to the inner: following this, they assume that compounds must contain a different constituent structure. However, Nespor and Vogel also state in the beginning of their 1987 analysis that the phonological component of language involves subsystems, including one dedicated to Lexical Phonology. If the intervocalic voice rule is blocked by an earlier constraint in the lexical specification for *girasole*, then the underlying representation is already specified for [s]. Simply put, there is no rule to be applied.[19]

The other argument for the Composite Group constituent hinges upon the rejection of recursivity within the prosodic hierarchical framework. The issue of recursivity in phonology is a contentious one: Pinker and Jackendoff (2005) rejected any and all possibility of recursion in phonological structure:

> We note that the segmental and syllabic aspect of phonological structure, though discretely infinite and hierarchically structured, is not technically recursive. Recursion consists of embedding a constituent in a constituent of the same type... this does not exist in phonological structure.

But there have been arguments in support of the existence of recursion in phonology, particularly in intonational structures (Ladd, 1986, 1996; Hunyadi, 2010). Evidence

[19]Baroni agues that intervocalic [s] voicing in Northern Italian is sensitive to the morphological structure of the words: "The frequency of [s] realizations seems to depend on how “transparent” the morphological constituency of a word is. At one extreme, nonce formations must always be treated as morphologically complex, since speakers can only parse them if they are able to decompose them into morphemes"(2001: 141). Indeeed, the example used here, the compound *girasole* (from the verb *girare* + noun sol) involves both semantic and morphological complexity.

from both empirical and experimental studies of language (e.g. van der Hulst, 2010 and Vigário, 2010) has led to the hypothesis that recursion may in fact exist in the higher structures of the prosodic constituency (that is, at and above the word level).

2.3.3 In Favour of A Recursive Phonological Word Analysis

The discussion above has focused on the Phonological Word constituent; we have seen that there are several linguistic items that raise questions about the structure of this constituent. Clitics pose problems because they often reduce and lose stress, making them incapable of meeting the requirements of a Phonological Word (cf. §2.3.1). Compounds, on the other hand, contain multiple Phonological Words yet simultaneously act as a single prosodic unit in connected speech (cf. §2.3.2). Nespor and Vogel (1986, 2007) and Hayes (1989) acknowledged the issues raised by these items and accordingly introduced a new constituent into the Prosodic Hierarchy designed to deal with them (the Clitic Group and, later, the Composite Group). However, as we have seen in both sections, these constituents are not without their own problems: the Clitic Group was abandoned due to the fact that it violated the SLH by dominating multiple Phonological Word constituents (cf. example 2.18). We also saw that the other proposed constituent, the Composite Group, ultimately requires a recursive structure despite being designed to eliminate it.

So it seems that we are left with two options here: either allow recursion at the Phonological Word level, or introduce new and specific constituents. §2.3.1 and §2.3.2 have illustrated the problems with the introduction of the Clitic and Composite Group constituents. In the remaining pages of this chapter, we will argue in favour of a recursive Phonological Word analysis. Support for this argument comes from two sources: experimental evidence from speech production studies and Selkirk's recent (1995, 1996, 2011) modifications to the Prosodic Phonology framework.

2.3.3.1 Experimental Evidence for the Prosodification of Clitics and Compounds

The first experimental study to elicit evidence for constituent structure in English came from a series of tasks designed to gather data on rapid movement sequences in speech and typewriting. While their intent was to study how groups of words were stored in short-term memory prior to speaking and/or typing, Sternberg et al. (1978) actually provided some of the first experimental evidence for the Phonological Word. Among the experiments they ran, two are particularly relevant to the Phonological Word constituent. In the first task, subjects were asked to store and produce both monosyllabic and disyllabic words. The disyllabic words were stressed on the initial syllable and used the monosyllabic words as their initial syllable (for example: *bay/baby, rum/rumble, track/tractor, cow/coward*).[20] They found that production latency increased linearly with list length in this task.

In a second experiment, they added unstressed function words (such as *and*) to the lists: results from this task indicated that function words did not alter the latency pattern in speakers. Sternberg et al. related this latency pattern to a buffer caused by "processing stages" in speaking: a "retrieval stage" (where latency depends on the number, but not the size, of units being retrieved) and an "unpacking stage" (in which latency depends on the number of constituents, not units) (1978: 148). Moreover, they were able to identify the stress group, a buffering unit that predicted latencies and durations in their experiments. Sternberg et al.'s isolation of this speech segment is crucial for our understanding of the shape of the Phonological Word constituent: the "stress group" that Sternberg and his colleagues isolated as the programming unit in their findings was in fact the Phonological Word.

Ferreira (1991) ran a series of experiments based on those of Sternberg et al.'s to study the effects of syntactic complexity on speech production. Instead of using lists, she used sentences of increasing Phonological Word length, syntactic complexity, or semantic plausibility to examine the planning process. The word length experiment contained four

[20]This was done in order to replicate the addition of an unstressed syllable as cleanly as possible.

sentences of varying length and structural complexity. Some examples include:

(2.31) **(a) Short condition**:
The river empties into the bay that borders the little town.

(b) Long condition:
The river that stopped flooding empties into the bay that borders the little town.

Her findings supported those of Sternberg et al. (1978).: namely, that initiation time was directly correlated with the number of Phonological Words in an utterance. The greater the total number of Phonological Words in the utterance, the longer it would take speakers to initiate speech. Crucially, she also noted that "the length effect... suggests further that the sentence production system attempts to produce a structurally defined unit-a unit that can contain varying numbers of words" (1991: 217).

In 1997, Wheeldon and Lahiri expanded upon Sternberg's paradigm and measured reaction times for cliticised utterances in a series of prepared speech tasks. They found evidence that the Dutch determiner *het* ("the") encliticised with the neighbouring word and formed a single prosodic word, resulting in similar naming latencies to sentences that contained no determiner. In these tasks, speakers were asked to respond to questions such as *Wat zoek je?* ("What do you seek?") using prompts from clitic and non-clitic conditions. The experimenters found that the preparation time for sentences containing clitics (e.g. *Ik zoek het water*, "I drink the water") were no different to those that contained none (e.g. *Ik zoek water*, "I seek water"). This indicated that clitics were attaching to the neighbouring word and forming a single prosodic unit, and that the number of prosodic words in both clitic and non-clitic sentences were the same.

Table 2.1: Results: Experiment 2, Wheeldon & Lahiri (1997)

Condition	Expected Response	Content Words	Phonological Words	Syllables	Mean rt (ms)
(1) Clitic	Ik zoek het water. *I seek the water.*	2	2: [ik zoek het] [water]	5	396
(2) Non-clitic	Ik zoek vers water. *I seek fresh water.*	3	3: [ik zoek] [vers] [water]	5	410
(3) Pronoun	Ik zoek het. *I seek it.*	1	2: [ik zoek] [het]	3	396
(4) Control	Ik zoek. *I seek water.*	1	1: [ik zoek]	2	377

This pattern held even if the first word was a full Phonological Word (e.g. *Riet zoekt het water*, "Riet seeks the water"). However, although the results suggested that function words were treated as weak elements in that they attached to a lexical word (and thus formed a single Phonological Word for the purpose of encoding), it was not evident which direction these weak words attach. For instance, they could be attaching rightwards to *water* rather than *zoek*. To confirm the direction of cliticisation, the experimenters presented an online naming task on the assumption that the size of the initial Phonological Word would predict naming latencies. If *het* was in fact encliticising leftwards, then the sentence containing the clitic would yield the longest latencies. As predicted, clitic sentences took significantly longer to produce in this task.

In summary, they found that neither the number of lexical words, content words, nor syllables in an utterance predicted naming latencies in clitic sentences in Dutch. Clitics attached leftwards to the initial phonological unit [*Ik zoek*], forming a single Phonological Word [*Ik zoek het*] which resulted in similar naming latencies to other sentences that contained the same number of Phonological Words, even if they contained fewer lexical or content words. The structures that predicted the naming latencies were always phonological in shape, no matter the type of task (online or delayed).

In 2002, Wheeldon and Lahiri used the same experiment to investigate the naming of compound words in Dutch. They found that noun-noun compounds (e.g. *ooglid*, "eyelid") elicited significantly shorter results than adjective-noun phrases (e.g. *oud lid*, "old member"), even though they shared the same number of lexical words. The latencies for noun-noun compounds were similar to monomorphemic words (e.g. *orgel*, "organ").

Table 2.2: Results: Wheeldon & Lahiri (2002)

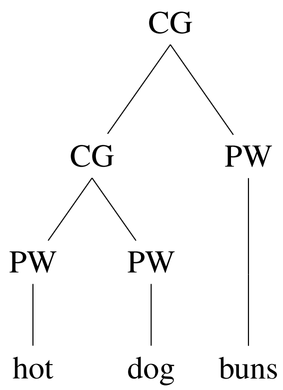
Conditions	**Example**	**Lexical Words**	**Phonological Words**	**Mean rt (ms)**
Compounds	*vliegtuig* (airplane)	2	1	349
Adjective-noun phrases	*fel tuig* (bright harness)	2	2	360
Simple initial stress	*farken* (pig)	1	1	348
Simple final stress	*fornuis* (stove)	1	1	351

Nespor & Vogel (2007) attributed the latencies in Wheeldon & Lahiri's (2002) experiment to Composite Group constituents: they dismissed the use of recursive prosodic structures and maintained that it was the Composite Group that predicted "structural complexity in language processing" (p. xx). But there are linguistic items whose structures demand recursive phonological structure, such as multiword compounds (e.g. *hot dog bun*)[21] and cliticised compounds (e.g. *hot dog buns are*).

The compound "hot dog bun", if submitted to the Composite Group (CG) constituent structure, would still require multiple, recursive prosodic structures, because *hot* and *dog* are individual Phonological Words. If Phonological Words cannot dominate other Phonological Words, then the dominating node for *hot dog* must be a Composite Group node. However, *bun* is a part of the compound structure, and the only way to illustrate this correctly is to include another Composite Group node dominating the entire structure as in example 2.32 below:

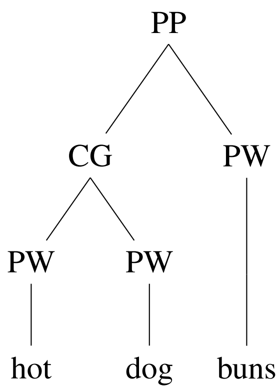
[21] Certainly in rapid AmE speech, "hot dog bun" will be pronounced as a single compound word: [hɑtdɒgbʌn].

(2.32) hot dog buns

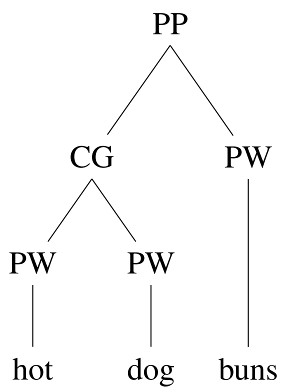


Treating *hot dog bun* as a Phonological Phrase is not ideal either, because we would be left with the same problem as in §2.3.2: there would be no way to distinguish between the compound *hot dog buns* (bread rolls that you put sausages into) from *hot dog buns* (a possible phrase to describe some warm dog buttocks in AmE!).

(2.33) bread rolls



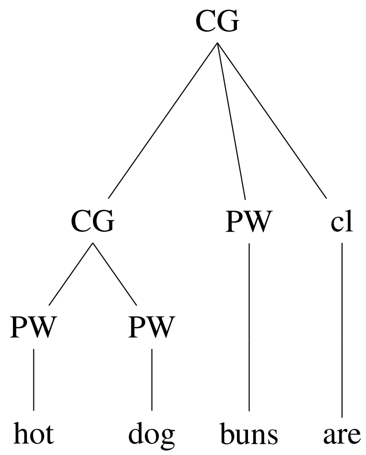
(2.34) warm dog buttocks



Furthermore, this structure violates the SLH by allowing *buns* to skip a level and attach directly to the Phonological Phrase.

Things only get worse for the Composite Group when a clitic is added into the mix, as in *hot dog buns are*, where the auxiliary *are* will often reduce and cliticise leftwards to a host word during normal speech. If we were to treat this structure as the Composite Group, it would still involve recursion (example 2.35):

(2.35) Hot dog buns are.



Again, reverting to treating "hot dog buns" as a Phonological Phrase would invoke another violation of the SLH, because a clitic cannot skip a level to attach to a Phonological Phrase. Looking at these complex prosodic structures above, we immediately see two things: the first is that these structures all seem to require either recursivity or a violation of the SLH despite being proposed as answers to these very problems. Second, the experimental evidence indicates that these items are being treated as single Phonological Word units: that is, clitics attached to lexical words to form single units, and compounds were also treated as single units. Taken together, these sources suggest that we cannot simply excuse recursivity in these structures. But what does this mean for the strict rules that form the foundations of Prosodic Hierarchy?

2.3.3.2 Modifying the Strict Layer Hypothesis

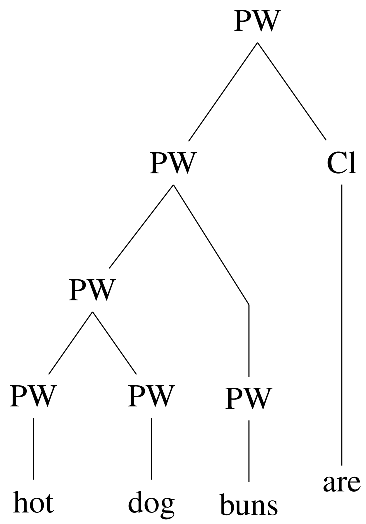
Selkirk addressed violations of the SLH (such as recursivity) in her 1995 paper, in which she also maintains that an additional prosodic constituent is not necessary to account for the behaviour of function words and compounds. These "primitive component restraints", drawing on the framework of Optimality Theory (cf. Prince and Smolensky, 1993), are listed below:

(2.36) Constraints: (Where C^n= some prosodic category)
1. Layeredness: No C^i dominates a C^j, j > i.
2. Headedness: Any C^i must dominate a C^{i-1} (except if $C^i = \sigma$).
3. Exhaustivity: No C^i immediately dominates a constituent C^j, j < i-1
4. Nonrecursivity: No Ci dominates C^j, j = i.

In the Optimality Theory framework, constraints on phonological representations are considered "violable". Optimality Theory allowed violations of constraints under two circumstances: if the output structure violates a constraint that is **higher** than the constraint violated in the original structure, or if the alternative output proposes **more** violations of the same constraint (1995: 209).[22] Therefore, the constraint of "nonrecursivity" can be violated if the only other options are to treat the dominating constituent as either a foot or a Phonological Phrase (Selkirk, 2011).

If we allow the SLH to be treated as a series of individual constraints on phonological structure, then the treatment of the complex structures above (*hot dog buns*, *hot dog buns are*) as recursive Phonological Word structures requires minimal adjustment of the constraints:

(2.37) Hot dog buns are.



[22] Selkirk stipulates that the constraints themselves are "universal", however "it is differences in the ranking of the relevant constraints that must be given responsibility for this cross-linguistic surface variety in prosodic structure." (1996: 209)

Recursion at the Phonological Word level fills both of the OT criteria discussed above: arguing for recursion at the Composite Group level would involve a violation higher in the constituency, and the alternative (allowing the clitic to attach at another level) would involve skipping levels. As to the issue regarding whether recursion exists in phonology: Selkirk (1995) argues that the recursive qualities of the phonological structures "mirror" the recursive morphosyntactic structure while "respecting" the Phonological Word boundary. Addressing the question in a different way, Scheer (2011) suggests that while the phonology itself is not recursive, the prosodic constituents themselves can exhibit recursivity.

2.4 Summary

The aim of this chapter was to provide a detailed review of a theory of pre-ordered constituents. In §2.2, we traced Prosodic Phonology from its earliest roots in the grammars and manuals of the eighteenth century to its current status. We focused on one constituent in particular: the Phonological Word, where syntax, morphology, and phonology interact the most. We saw early glimpses of the Phonological Word's makeup in Steele's (1779) cadences (in particular, the cadence structure for "sentence in"), where it was treated as the domain of syllabification, and Liberman and Prince's (1977) "mot" level where it is the bearer of prominence in multiword structures. In discussing the structure of this constituent, we examined complex morphosyntactic items such as compounds and clitics which have been the source of much debate. In the end, it seems that we are left with two options for dealing with these items: either the addition of a new constituent into the prosodic hierarchy, or there ordering of the SLH into a series of violable constraints.

We argue that a recursive treatment of these items is the best approach for several reasons.First, it seems that recursivity is difficult to avoid, particularly when dealing with cliticised compounds such as "hot dog buns are". We saw that the Composite Group, despite being introduced to avoid recursion, requires it nevertheless with these complex structures. Secondly, by adopting an OT-style approach to the SLH, we have been provided an elegant way of coping with the complexities of compounds and cliticised structures within the current framework of prosodic constituent theory. The theory is not weakened by being having the SLH fleshed out into violable constraints; rather, it is enriched. Taking into account the empirical and experimental evidence discussed in this section, it seems that we must allow for some recursive behaviour at the Phonological Word level. In the following chapter, we will place theory in reality, looking at where the subsystem that prepares prosodic constituents may find its place within the cognitive processes that plan and arrange speech for articulation.

3

Phonology in Speech Production

3.1 Overview

Just as the language is seen as being composed of a series of generative processes in theoretical frameworks, so is the cognitive architecture that makes up its physical reality. Models agree that the production of speech involves a series of stages that prepare intermediate representations at each step: semantic representations are prepared during conceptualisation, grammatical representations during lexicalisation, and phonological representations during word-form encoding. However, only recently have we started to address what happens between these stages and, in particular, what happens when grammatical representations exit the lexicalisation stage and enter the phonological encoding stage. In this chapter, we will introduce language production modelling and focus on the evolution of the phonological encoding stage from a simple "large-scale" stage in charge of attaching phonology to syntax, to a series of subprocesses that infuse prosodic representations with phonological features (Poeppel & Hickock, 2004).

This chapter will present a historical analysis of language production modelling, with a special focus on the processes that prepare post-lexical representations for speech. In §3.2, we will review how early analyses of speech errors have contributed to our current understanding of a multi-stage process of preparing spoken language. Additionally we will highlight how early models that featured discrete "steps" and representations led to the consideration of different processes to deal with specific prosodic and phonological features. Each of the models discussed here has contributed valuable information about production processes and the structure of the mental representation of language. Two models in particular, those of Levelt (1989) and Roelofs (1997), are notable for the attention they pay to the phonological encoding process. In our analyses, we will look closely at how these models treat the process that prepares prosody in spoken language. §3.2 is devoted to the discussion of non-native language production models, and what they reveal about the phonological encoding process in L2 speakers specifically. Again reference is made to the models of Levelt and Roelofs, particularly in respect to the features of non-native spoken language.

3.2 Language Production in Native Speakers

3.2.1 The First Models: The Speech Error Tradition

The study of language production has its roots in a rich history of both psychological and linguistic theory. In his publication *A History of Psycholinguistics: The Pre-Chomskyan Era* (2013), Levelt attributes much of our current understanding of the speaking process to late nineteenth studies of psychology (e.g. Wernicke, 1874) and historical language change (e.g. Sturtevant, 1917).[1] Around the same time as Sweet and

[1]This publication provides what is arguably the most comprehensive survey of pre-SPE language modelling to date. Levelt traces, in great detail, the historical roots of modern psycholinguistics, giving equal footing to comparative linguistics, speech impairment, child development, and laboratory research. *A History of Psycholinguistics* is an essential reference for any researcher wishing to embark on a psycholinguistic study.

Sievers were preparing their grammars, diagrams of the presumed architecture of spoken language began to appear, e.g. Baginsky (1871), Wernicke (1874), Kussmaul (1877), and Lichtheim (1885). These diagrams were all formed with regard to the anatomical localisation of speech, in order to isolate specific speech impairments such as aphasia. Consequently, the diagrams were very detailed, often consisting of several layers and levels.[2] Out of this tradition of diagramming came the treatment of language not as a simple large-scale operation, but as a multi-stage process with dedicated stages. In 1877, Kussmaul had already assumed three different stages for the *Act der Sprechens* ("act of speech"), remarkably similar at first glance to the three-step models we have today:

> *Somit zerfällt der Act des Sprechens stets in drei Stadien oder Vorgänge: 1) Die Vorbereitung der Rede in Geist und Gemüth; 2) die Diction oder die Bildung der inneren Worte sammt ihrer Syntax; 3) die Articulation oder die Bildung der äusseren Worte oder „Wörter" unbekümmert um ihren Zusammenhang in der Rede.* (1877: 14)[3]

In his 1880 study of language history, Hermann Paul noted that speech sounds were often lost or displaced in similar or adjacent elements, and surmised that certain aspects of language change could be linked to certain types of speech errors (Levelt, 2013: 156). Large collections of speech errors began to be compiled during this time, the most notable by Rudolf Meringer and Carl Mayer; their 1895 publication of *Versprechen und Verlesen* contained over 8,000 spoken and written errors from regular, everyday communication. Like Paul (1880), Meringer and Mayer found that speech errors were predominantly errors of ordering, in which certain segments of words were incorrectly displaced into functionally similar positions. They argued that these displacements were not random; rather, most frequent types of errors could be attributed to *eines gewissen geistigen Mechanismus* ("a particular mental mechanism"). They also found that many errors

[2]Cf. the Wernicke-Lichtheim model (Lichtheim, 1885), which was certainly one of the more complex architectures of speech and accompanying deficits. The diagram was built around centres responsible for auditory, motor, conceptual, reading, and writing systems.

[3]Thus, the act of speech always falls into three stages or processes: 1) The preparation of the utterance in the mind, 2) the diction or the formation of the internal words together with their syntax; 3) the articulation or the formation of external words or "words" unconcerned by their relationship/connection (i.e. semantics) in the utterance.

corresponded well to specific units of grammar (features, phonemes, and morphemes).[4] The cataloguing and analysis of similar groups of errors led researchers to recognise that errors occurred at certain 'joints': crucial points where units were connected together (Bock, 1996). Based on the locations of these joints (and the patterns associated with them), they (and other researchers) were able to begin building models of the production of human speech. Speech errors remained the primary source of data for the modelling of speech into the twentieth century.

3.2.2 Twentieth-Century Models of Speech Production

Production models of the twentieth century fall into two categories: serial and interactive. Both offer the same basic processes (message generation, access and assignment of grammatical and phonological information, and preparation of articulation, but they differ in how the levels interact with one another. Serial models (Fromkin, 1971; Garrett, 1976b, 1976a; Levelt, 1989) assumed that the production of speech occurs in a strict, cascading order, and that all of the processes are independent from one another. Interactive models (MacKay, 1982; Harley, 1984; Dell, 1986) assume that processes work simultaneously and information feeds back through the model after encoding. This distinction between the models will become significant when we focus on the phonological encoding stage in particular. Table 3.1 presents the major models of language production we will be discussing.

Victoria Fromkin (1971) produced one of the first complete models of speech performance with the aim of isolating the underlying representations of the processes. In line with the classical three-stage models, her model also had separate, distinct processes for generating messages, lexical information, and phonological segments.[5] She saw language

[4] *"... die Laute eines Wortes, eines Satzes, und auch die Worte untereinander in ganz eigentümlicher Weise verbunden und verknüpft sind"*, ... the sounds of a word, a sentence, and even the words themselves are all interconnected and linked in a peculiar way." (Meringer & Meyer, 1895: 10)

[5] "What is apparent..." she writes, "is that, despite the semi-continuous nature of the speech signal, there are discrete units at some level of performance which can be substituted, omitted, transposed, or added... "(1971: 30).

Table 3.1: Major Models of Language Production

Model	Date	Type	Data Used	Encoding of Phonological Material
Fromkin	1971, 1973	serial	speech errors	- noted that slips of the tongue "suggest that speakers recall the number of syllables" in a word. - isolated "strings of phonological segments". -maintained that models of speech production had to take into account the features, segments and syllables that constituted planning units.
Garrett	1976	serial	speech errors	- isolated two levels of processing: a level where syntactic functions were assigned, and a level where forms were organised.
Shattuck-Hufnagel "Scan-Copier"	1979	serial	speech errors experimental data	- sets of ordered lexical units are associated to independent syllable frames during phonological encoding -three levels of processing: morphological/metrical spellout, segmental spellout, and phonetic spellout
Butterworth	1980	serial	speech errors	-'phonological lexicon'
Harley	1984	interactive	speech errors	-'Phonological Buffer'
Dell "Spreading Activation"	1986, 1988	interactive	speech errors	- based on the mechanics of spreading activation - defines units of phonological encoding as morphemes, syllables, rhymes, segment clusters, segments and features - nodes represent the units and form a **hierarchical structure**; activation spreads to all related nodes.
Levelt	1989, etc.	serial	experimental data	-a **prosody generator** takes metrical, phrasal and syntactic information and generates prosodic units during encoding.
Roelofs	1997, 1999	interactive	-experimental data and speech errors	-semantic representations activate nodes in the syntactic and phonological networks. -spreading activation and prosody generator.

production as a fluid process involving the preparation and assignment of features during different stages. In particular, she viewed the output of the grammatical preparation stage as "a syntactic structure with semantic and syntactic features specified for the word slots" (239). In Fromkin's model, phonological features are assigned at a later stage, when phonological rules dictate how the strings of phonological segments are arranged. Crucially, they are arranged "with the syntactic bracketing remaining intact" (240). In 1973, Fromkin published an additional catalogue of nearly 600 speech errors. In the accompanying analysis, she emphasised the importance of evidentiary data for each level of a speech performance model, and maintained that speech errors were able to provide invaluable insights to the nature of language behaviour and speech production. Fromkin's model laid the framework for our understanding of the individual processes occurring in language production, and served to establish a specific stage where phonological features were assigned to grammatical representations.

In 1976, Merrill Garrett proposed a model which focused heavily on how semantic material was generated during speech. In his model, he made a strict distinction between two processes he referred to as *language production* and *sentence production*: *language production* was responsible for the generation of messages and concepts (i.e., the nonlinguistic aspects of speech), while *sentence production* was the process by which structural units of speech were gradually "translated" into spoken utterances. The levels in Garrett's model acted independently from one another, each accessing its own information from the associated store to prepare its own representations. Garrett maintained that the autonomy of each stage was supported by the behaviour of speech errors: like his predecessors, he noted how phoneme exchanges (e.g. *hack rat* for *hat rack*) often affected segments occurring in similar environments, while word exchanges tended to retain grammatical functions. These findings led him to view production as a series of as "quasi-independent levels" which reflected corresponding speech error types (see Figure 3.1).

Figure 3.1: The image originally presented here cannot be made freely available via ORA because of copyright. The image was sourced from Garrett, 1976: 176.

"Sentence production" was confined to the Functional and Positional Levels in his model. At these levels, frames were built according to syntactic features and grammatical representations; then phonemically-specified lexical representations were "inserted" into the frames (176). This process was similar to that of Fromkin's model, where phonological features were assigned to pre-arranged lexemes.

A notable feature of Garrett's model was the treatment of function words: he hypothesised that function words were not accessed until the Positional Level, while content words became available earlier at the Functional Level. This underlines that he saw a mismatch in the spoken features of language, such that he assigned them to a different level. While certain aspects of the production stages were more defined in Garrett's model, we begin to see issues here with the strict association of speech errors to processes. In particular, the model encountered difficulty when it attempted to account for speech errors that involved several levels, such as sentence and phrase blends (e.g. *take a tab* for *take a cab*). Such an error arguably occurs both at message (the activation of *tab* and *cab*) and form levels, something that should be impossible in Garrett's model.

As part of his analysis, Garrett worked with Stefanie Shattuck-Hufnagel to amass nearly 6,000 speech errors in the "MIT-CU Speech-Error Corpus". Analysis of this data led Shattuck-Hufnagel to identify particular behaviours of speech errors; as a result, she devised constraints on the behaviour of speech errors in her model. Like Garrett and Fromkin's models, Shattuck-Hufnagel's (1979) "Scan-Copier" model associated patterns of speech errors to different processing units and levels during production. However, it also proposed that segments were accessed from a long-term lexical store and placed into a short-term memory store, where they were selected for the correct position in a frame.

The Scan-Copier model contained three mechanisms: (1) a representation composed of serially-ordered slots and matching target concepts, (2) a scan-copier mechanism, which chose the correct planning segment and inserted it into the matching slot, and (3) two separate monitors which checked for correct slots and errors.

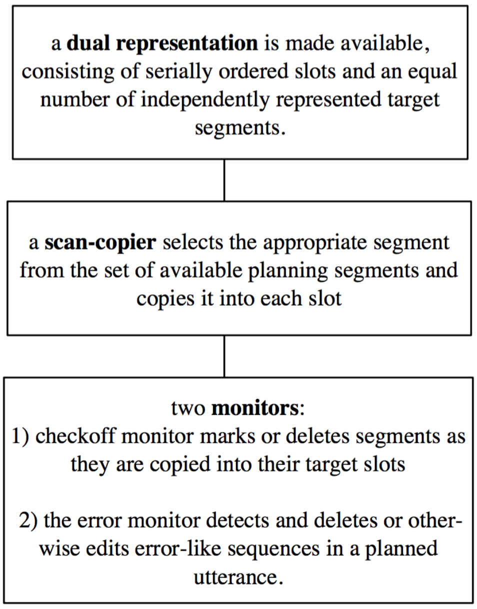


Figure 3.2: Shattuck-Hufnagel's (1979) Model

Notably, Shattuck-Hufnagel's model was the first to address prosodic structure in language production: the prosodic structure of an utterance (also stored in the long-term memory) indicated what segments the scanning mechanism needed to fill each slot in the frame. In the Scan-Copier model, the process of arranging lexical representations into frames during the phonological encoding stage became significantly more complex: rather than simply assigning abstract phonological information to frames containing existing lexical representations, the model filled them with segments that were predetermined by stored prosodic structures. With this model, we see not only the content within the frames becoming significantly more detailed, but also the structure of the frames themselves.[6]

This was a turning point in the modelling tradition, and it brought about a number of models in which the previously-underrepresented phonological encoding stage finally received attention. In his "Full-Listing Model", Butterworth (1980, 1983) proposed a "phonological lexicon". After a production mechanism accessed both the semantic and the phonological lexicons, it used a phonological "editor" to filter out any mismatch. Harley (1984) also isolated phonological operations to one level: the output of the lexicon was run through a "Phonological Buffer" in his model, at which point all phonological information was assigned to the representation. Stemberger (1985) looked at the shape of phonological units in the lexicon, using a wide variety of speech errors as evidence: in addition to regular speech errors, he also used errors from aphasia and child language. The mid-1980s also saw the rise of a different type of modelling: connectionist modelling (Stemberger, 1985; Dell, 1985, 1986; MacKay, 1987). One of the motivating factors in the introduction of the connectionist model was related to the restrictive nature of the slots-and-filler models (e.g. the Scan-Copier Model). Connectionist modelling instead proposed that speech production involves a cooperative and interactive network, in which activation 'cascades' over nodes and arcs (Stemberger, 1982; Dell, 1985, 1986; MacKay, 1987). Segments are selected based on their level of activation; in this way, the different stages of production interact with one another as activation spreads through them.

In one of the most well-known connectionist models to date, Gary S. Dell (1986) presented a design that followed the basic tenets of previous models, while at the same time allowed for multi-directional activation of information. In his model, Dell saw production as a series of processes that were governed by generative rules (syntactic, morphological, and phonological). These rules dictated how nodes were activated and formed specialised frames that contained categorised slots at every level. Semantic nodes

[6]The complexity of this process, and indeed of the entire model, was also one of its major limitations: Levelt (1992) found the process of generating an empty word skeleton from a stored phonological representation, filling it with segments, and then restructuring it to fit the requirements of the utterance unnecessarily "wasteful" for the speaker.

spread their activation to lemma nodes, then on to phoneme nodes. At the phonological level, activation made two types of information available: (1) the constituent phonemes of the representation, and (2) the word's frame.

Figure 3.3: The image originally presented here cannot be made freely available via ORA because of copyright. The image was sourced from Dell, 1986: 290).

In Figure 3.3 above, Dell's model is activating the morphosyntactic and phonological information for the utterance, "some swimmers sink". The most highly-activated nodes spread their activation from the lexical network down into the phonological encoding stage. After syntactic framing and grammatical features are prepared, the processor goes on to activate a syllabic structure (seen in the tree structures that accompany the diagram in Figure 3.3). In this sense, the model retained Shattuck-Hufnagel's slots-and-fillers approach to phonological encoding: frames are built according to prosodic information (e.g. number of syllables or phonemes). However, as opposed to the earlier models, the frames in Dell's model consisted of "inter-connected set(s) of nodes, each standing for some structural category such as onset constituent" (Dell et al., 1993: 150). Frames were filled with the most highly-activated lexical nodes, but they were built according to the specific phonological rules of the language. Crucially, the building and filling processes were seen as distinct processes in this model.

In the models discussed in this section, we have seen the evolution of the phonological encoding stage move from a large-scale stage where phonology was simply applied to existing lexical representations, to a multi-faceted stage in which the retrieval and generation of phonological frames was seen as a separate process to the spoken features of the word (that is, the word's phonological content). However, while these models were invaluable to the interpretation of speech production, they all had one significant, shared weakness: their reliance on speech-error evidence.

The most obvious problem with speech-error data is observer bias in collection. The speech-error corpora from the 1970s and 1980s contain data that was written down during day-to-day activities, usually by one observer, and unsubstantiated by other observers. But, as Meyer (1992) notes, "errors of stress, intonation, or phonetic encoding" were rarely recorded in collections such as Fromkin (1975, 1977) or Cutler (1979). This, she muses, could be happening for one of two reasons: either speakers do not regularly make these types of errors, or they go unheard by observers. Experimental evidence points to the latter: in a number of experimental tasks, listeners repeatedly failed to pick up violations of phonetic rules (Warren et al., 1970; Marslen-Wilson & Welsh, 1978; and Cutler, 1981). Another problem relates to the limitations of speech error data: although speech errors correspond well to individual features and units, they do not reveal much about larger units of speech (such as phrases and clauses). Even the multi-dimensional connectionist models focused primarily upon the encoding of single-word utterances or lists of unrelated words.

Consequently, researchers began to explore other sources of data in order to reveal information about previously underreported areas in the production process. Experimental methods had already been used to elicit results about certain linguistic processes in the 1960s: Oldfield and Wingfield (1965) used reaction time experiments to study word frequency effect, Brown and McNeill (1966) were able to induce ToT (Tip-of-Tongue) errors in their experiments, and Baars et al. (1975) studied speech errors by having participants read lists of words at increased speed.[7] Experimental results were often able to reveal significant information about previously underreported areas, particularly in phonological encoding (such as syllabification and prosodification). Furthermore, studies on language impairment were revealing significant findings regarding language production stages (e.g. Butterworth, 1989). It soon became clear that the existing models had to be adapted to be able to account for this new data.

3.2.3 Levelt, WEAVER, and WEAVER++

Willem Levelt's (1989) model introduced a serial model that focused heavily on the requirements of connected speech. In his model, the production of fluent speech was carried out by three discrete processing components: (1) the Conceptualiser, (2) the Formulator, and (3) the Articulator. The Formulator was responsible for infusing conceptual representations with lexical shape, as well as preparing the phonological features of the utterance. It begins by accessing the item's lexical form, then building a phonological representation from information provided in that form. When the preverbal

[7] Of course, chronometric experiments on language production had been conducted far earlier: Cattell (1885) used picture-word and list-naming tasks to elicit speech errors.

message enters the Formulator, it gains access to an existing repository of information about a language's semantics, syntax, and words: this repository is known as "the mental lexicon". Entries in the mental lexicon, called "lemmas", become available prior to syntactic construction but are entirely unspecified for phonological form. These lemmas are not necessarily word-sized; Levelt argues that transposition errors (e.g. *wish a brush* for *with a brush*) suggest that lemmas can also be constituents smaller than words. Accordingly, the Formulator selects the lemma most suited to the concept and retrieves its accompanying syntactic information (much like the slots-and-fillers models). The output of grammatical encoding is a "surface structure" that is arranged hierarchically, where lemmas act as terminal nodes. This output then enters the phonological encoding stage (cf. Figure 3.4).

The goal of the phonological encoding mechanism is to "generat[e] a string of syllables that the Articulator can accept and pronounce" (Levelt, 1989: 318). At this stage, the Formulator accesses the item's "lexeme" (or "lexical form") from the mental lexicon; every lemma has a corresponding lexeme that contains its phonological specification.[8] Because the shape of a word depends on its environment, the phonological specification is still abstract at this stage. Along with the abstract phonological specification, the Formulator accesses basic metrical information stored for the lemma: this includes the number of syllables it contains, the weight of these syllables, and possibly even the default metrical frame. However, it does not include the syllables themselves. This is one of the key stages where Levelt's model differs from others: he maintains that 'default' syllable structure is **not** stored along with the other phonological information for each lemma.[9] Instead, it is retrieved from the mental lexicon as sets of sublexical and subsyllabic units

[8]Originally described by William James in 1890 as "a gap that is intensively active", the Tip-of-the-Tongue (TOT) phenomenon has long been referenced in studies of speech errors (cf. Loftus & Yuille, 1984 and Levelt, 2013). This error is a common one: the speaker knows the word they wish to use, but cannot retrieve it from memory. Often speakers can access some of the phonemes of the word, as well as the number of syllables and location of primary stress (Brown & McNeill, 1966). Levelt relates the TOT phenomenon to the blocking of successful retrieval of a word's lexeme.

[9]The prevailing argument here is that computing the default syllables during grammatical encoding and then re-computing them again during phonological encoding would be repetitive and unnecessary.

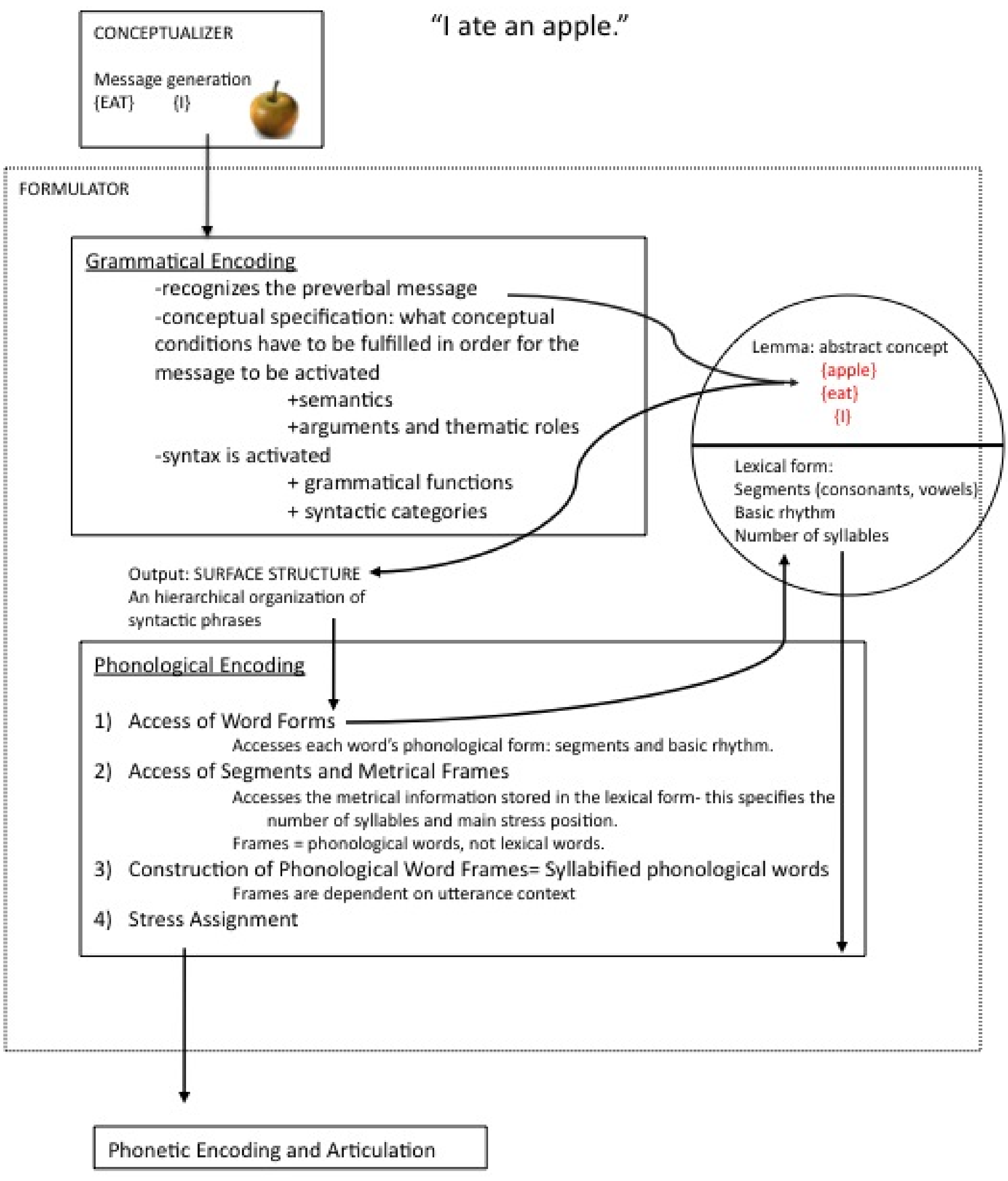


Figure 3.4: Levelt's (1989) Model

which are then positioned in independent structures (such as word and syllable skeletons) during phonological encoding (Levelt et al., 1999). Levelt's evidence comes from the fact that syllables can differ depending on their environment:

Table 3.2: Syllabification (Levelt, 1999)

Utterance	**Syllabification**
They are **selecting** me.	/sɪ/ /lɛk/ /tɪŋ/
They will **select us**.	/sɪ/ /lɛk/ /tʌs/
They **select me**.	/sɪ/ /lɛkt/ /mi/

For example (see Table 3.2), the word *select* occurs in three different conditions: the /t/ in *select* can act as a syllable onset (*selecting*, *select us*) or a syllable offset (*select me*). This process is not fixed, but instead occurs incrementally, online, and according to the syllables' environment. The process of syllabification allows for the representation of units in phonological terms: in this way, the sentence is arranged according to phonological units. The output of this process is the phonetic plan, which is then prepared for articulation in the final stage.

In addition to the incorporation of experimental evidence into his model and his aim to model features of natural, connected speech, another attractive feature of Levelt's model is its focus on the prosodification process. Specifically, he proposed that a distinct subprocess ("the Prosody Generator") is in charge of building templates for syllable constituents (onset, nucleus, coda) and segments. The input to the Prosody Generator consists of four types of information: the surface structure, metrical frames ("citation metrical spellout"), phonological segments ("citation segmental spellout"), and intonational meaning. Crucially, the Prosody Generator in Levelt's model did not build its frames according to the lexical output of the previous stage: it built them according

Further evidence against the storage of default syllables in the phonological code of the morpheme can be found in studies involving syllable-priming tasks (e.g. Ferrand et al. (1996), who found that participants named targets faster when they were preceded by a masked prime that was related to the first syllable of the target).

to **Phonological Word structures**. This was a significant departure from earlier models and changed the way that the phonological encoding process was viewed.

The WEAVER (Word-form Encoding by Activation and VERification) model was created by Ardi Roelofs in order to approach some areas that were unclear in models built exclusively on speech error data. WEAVER adopted aspects from Dell's (1986) model (particularly, the process of form retrieval by spreading activation) and Levelt's (1989, 1992) models (his approach to syllabification). It was the first model to attempt to account for reaction time data as well as the first to attempt a simulation of the process by computer algorithm. The model focused on the processes engaged in Levelt's 'Formulator' stage and, in particular, the encoding of phonological forms and features during production.

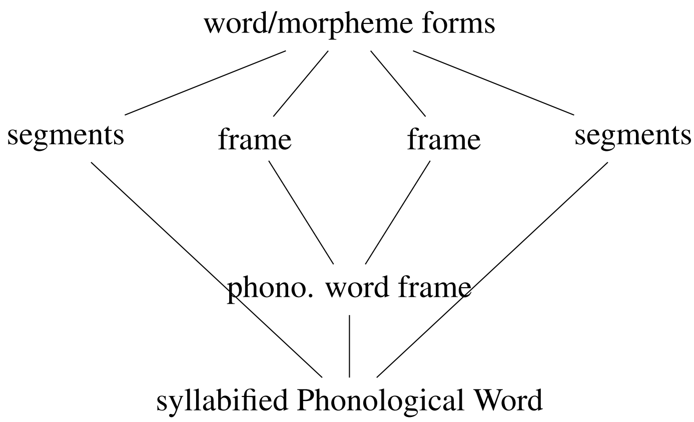
Instead of consisting of a one-way serial process, phonological encoding is achieved in WEAVER by a spreading-activation network in which roots and affixes are seen as nodes. Each level accepts a prepared representation as input and generates a new representation as its output. Post-lexical representations are generated from left to right, incrementally, according to the amount of activation they receive from linguistic rules and restraints.[10] Rather than conceiving this process as conducted by a central agent or generator, WEAVER proposes a team of procedures working in parallel to activate different parts of the word. Roelofs likened it to "several spiders making a single web" (1997: 250). One of the key issues that WEAVER addressed was that of flexible syllabification (i.e. when morphemes or segments change syllable position), a prominent feature of connected speech and polymorphemic word planning. Roelofs specifically references clitics here, whose production he proposes involves the creation of a "new Phonological Word" (1997: 253). In the model, these items were dealt with by a "binding mechanism", a process responsible for flexible syllabification in connected

[10] Roelofs' formula for spreading activation:
"$a(k, t + \Delta t) = a(k, t)(1 - d) + \sum_n ra(n, a)$ where *a,t* is the activation of node *k* at a point in time *t*, d is a decay rate ($0 < d < 1$) and Δt is the duration of a time step (in ms). The rightmost term denotes the amount of activation that k receives between t and $t + \Delta t$, where $a(n, t)$ is the output of neighbor n (equal to its level of activation). The factor r indicates the spreading rate" (Roelofs, 1997: 277).

speech and polymorphemic words.

In 1999, Levelt, Roelofs, and Meyer revised and expanded on WEAVER, resulting in a comprehensive word-form encoding model named "WEAVER++". In WEAVER++, Levelt's (1989) proposal regarding the shape of word frames is retained; further, they state that speakers do not generate lexical words at all during this stage: word forms activate a word's structural and segmental components separately based on contextual information for the utterance, and the structural frame which is generated is a Phonological Word frame (see example 3.1). In this model, individual phonemic segments of the activated morphemes are prosodified from the beginning of the word to the end. Only the phonological representations associated with the intended word is activated.

(3.1) Phonological Encoding in WEAVER++ (Levelt et al. 1999: 20)



Levelt et al. refer specifically to clitics at this point; they argue that the behaviour of these items (in that they attach left or rightwards in normal, connected speech) provides evidence for the construction of word frames that exceed the size of a lexical word (1999: 20). Although WEAVER++ does not dismiss speech-error data (it should, Roelofs points out, be able to account for any data), it takes the bulk of its evidence from experimental tasks. It is also able to illustrate features of normal, connected speech and effectively assign speech errors to their assumed levels. For these reasons, WEAVER++ appears the most capable model for dealing with experimental evidence about the phonological encoding process.

3.2.4 Interim Summary

The last forty years have seen a significant advancement in our understanding of the process that prepares the phonological features of spoken language. In the course of this discussion, we have seen that models have had to be adapted from ones built on speech error data to ones that can account for all features related to normal, connected speech.[11] In the following section, we will discuss how these models have affected the study of non-native language production.

3.3 Language Production in L2 Speakers

3.3.1 Overview

If the modelling of speech in native (L1) speakers is considered to be in its relative adolescence, then the equivalent for production modelling of non-native (L2) speech is still in its infancy. The primary focus of non-native speech studies has been on language learning and proficiency rather than the structure and generation of language itself. Reasons for this are manyfold: Palij & Aaronson (1992) cite both poor access to adequate speaker groups and a lack of general interest in non-native processing as contributing factors. As such, dedicated approaches to L2 speech production did not appear until the 1960s and 1970s, when researchers such as Kolers (1963) and McCormack (1974; 1977) began to question the structure of semantic stores in bilingual speakers.

[11] Does this mean that speech error data is now worthless to production research? Not at all. In an LSA Institute Workshop panel discussion in 2007, major production researchers (Dell, Schütze, Ferreira, Jaeger, Shattuck-Hufnagel and Stemberger, amongst others) considered whether the speech-error corpora were still useful in current language production research. They concluded that although the existing corpora did contain biased and sometimes ambiguous data, all methods of data collection is "flawed in one way or another" and that the data is not useless (Schütze & Ferreira, 2007: 391).

3.3.2 L2 Production Models

The first attempts at modelling non-native production focused primarily on the semantic and lexical levels of production. As in the early native models, researchers of L2 production used speech errors to isolate and argue for mechanisms dedicated to handling L2 phenomena. Macnamara (1967) proposed a process that allowed the speaker to 'switch' a language on and off ("The Switch Model"), Lipski (1978) envisioned a monitor that checked L2 constituents against existing L1 syntactic structure ("The Comparator Model"), and Albert et al. (1978) argued for a discrete process that checked for language-specific cues ("The Monitor System"). These early attempts at modelling L2 lexicalisation all shared one important feature: they introduced new stages or operations into the production process.

In the 1980s, theorists began to argue for dual-coding models containing two discrete systems: one that handled non-verbal objects or tasks, and one that handled linguistic information (e.g., Paivio & Desrochers, 1980; Potter et al., 1984). In 1992, de Groot ran a series of word translation tasks, primed lexical decision tasks, word association and semantic-relationship experiments, in an attempt to identify the role of the linguistic processing system in the bilingual memory model. As a result, hierarchical bilingual production models began to appear, mainly based on existing L1 models (cf. de Bot, 1992; Kroll, 1993; MacKay & Miller, 1994; Poulisse, 1997). Myers-Scotton (1992) used Levelt's (1989) model as a framework for her "matrix language frame model", which focused on the individual constraints involved with code-switching in multilingual speakers. Again, the attention paid to lexical encoding in these early models is hard to miss: De Bot (1992, 2004), Colomé (2001), Costa, Hiej & Navarrete (2006), Kroll et al. (2006), Green & Abutalebi (2008), and Simmonds et al. (2011) all offered hypotheses about how speakers select and encode non-native lexical structures during production. The theories all agreed that separate representations were activated during lexicalisation; from this point, however, the theories diverged. Kroll & Tokowicz (2005) cite a failure "to consider distinctions among levels of representation" in the production process as one of the major roots of this divergence (531).

Data elicited through experimental methods has aided in the comprehension of such levels and distinctions; evidence from numerous psycholinguistic tasks has indicated explicit competition between L1 and L2 lemmas in non-native speakers.[12] Picture-naming experiments conducted in a wide variety of languages have shown a delay in naming in the L2 language (e.g. Hanulová et al., 2011). Age of acquisition, word frequency, and inhibition mechanisms have all been investigated as sources of this delay; however it has not been attributed to any one production process.[13] As such, L2 production models have varied greatly on this matter: some assume an entirely dual system (with separate rules, sets of phonemes, and words for each language) while others present an extended system, in which the L2 interacts with L1 at any number of stages during production.

Pouliise and Bongaerts (1994) explained the delayed naming effect on the basis that L1 lexical items receive a higher level of activation: that is, the less fluent the speaker is, the less activation L2 lexical items receive during lexicalisation. In their three-stage model, based on Levelt's (1989) model, they propose that lexical representations from both languages are activated (Figure 3.5):

Figure 3.5: The image originally presented here cannot be made freely available via ORA because of copyright. The image was sourced from Poulisse & Bongaerts, 1994: 41)

[12] Cf. de Bot & Schreuder, 1993; Roelofs, 2003; Kroll et al., 2006; Finkbeiner et al., 2006; Bialystock et al., 2009; Green & Abutalebi, 2013.

[13] It is also possible that not one but several mechanisms are responsible for the delay (cf. Roelofs, 2003).

In the less-fluent speaker, slowdowns and disfluencies occur because the English lemma receives less activation than the Dutch.[14] The activated words are arranged in the syntactic frames of the intended (L2) language, then infused with phonological features from a shared phonological system.

De Bot (1992, 2004) also used Levelt's (1989) model as the framework for his "Multilingual Processing Model". His treatment of phonological encoding followed Levelt's quite closely, using adaptations only when necessary. De Bot theorised that, for each stage of production, there were three (associated) stores of information: conceptual features, syntactic and grammatical features, and word-form elements (2004: 12). Each node controlled the operations at each stage, and the cue to use a specific language continued to be relayed from the conceptual stage down to the phonological encoding stage. The language node associated with phonological encoding activates the phonological features associated with the intended language. However, because these features are selected from a shared phonological store, overlapping segments between the L1 and L2 representations can result in faster naming latencies (explaining such phonological priming effects as those found in Woutersen, 1997) (de Bot, 2004).

Green (1998) argued that the goal of speaking in L2 is stored in short-term memory: when the participant begins encoding the utterance, the message generator selects the lexical concepts that are required and this is flagged by the short-term memory mechanism to indicate the target language. In Green's model, both the L1 lemma and the L2 lemma are activated; however, the rule specifying L2 use will repress the non-target L1 lemma. When the phonological encoding stage begins, only the target (L2) phonemic segments are made available.

As we saw in the evolution of the native models, the introduction of experimental methods aided in revealing considerable evidence about how language is generated in non-native speakers. While the earlier models all attempted to explain the naming delay

[14] The relation of activation to fluency was first seen in de Groot (1992), who theorised that the L1 lemmas have a higher frequency than L2 and therefore are easier to access in less-fluent speakers.

by introducing new operations or mechanisms into the production process, these later models were able to adapt current native speaking (L1) models for the needs of the L2 speaker. Hierarchical, three-stage modelling allowed for further distinctions to be made, and for language-specific processes to be assigned to each stage. However, as with the native models, phonological encoding was again neglected in these models in favour of focusing on lexical and semantic processing in L2 speakers. What was really needed was a model that addressed phonological encoding in non-native speakers.

3.3.3 WEAVER++ for L2 Speakers

In 2003 (and 2006, with Verhoef), Roelofs produced an adapted version of WEAVER++ for non-native speakers. Citing extensive experimental evidence for the dual activation of phonological representations in both L1 and L2, he argued that phoneme representations, at the very least, were certainly shared between languages. In implementing this into the WEAVER++ model, Roelofs and Verhoef (2006) found that, with a few adjustments, WEAVER++ applied easily to non-native phonological encoding.

Recall that one of the conditions of WEAVER++ is that only the phonological representations linked to a particular lemma are activated during word-form encoding (Levelt et al. (1999) refer to this as the "discreteness assumption"). This condition explicitly disallows for cascading activation of non-activated lemmas from lexicalisation level to word-form level (Figure 3.6):

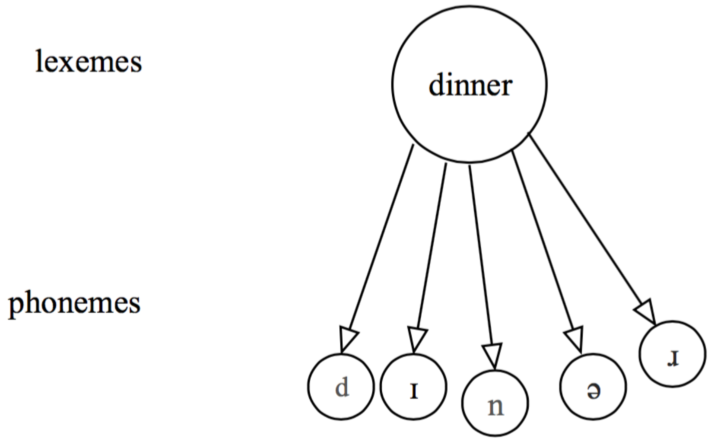


Figure 3.6: WEAVER++ for Monolingual Speakers

In Roelofs and Verhoef (2006), they loosen this requirement, allowing the model to account for the sharing of phonemes between the target (L2) and non-target (L1) languages. In Figure 3.7 below, both the English and French lexemes for the word "dinner" are activated, and we see how the sharing of phonemes facilitates the naming of *dinner* in English.

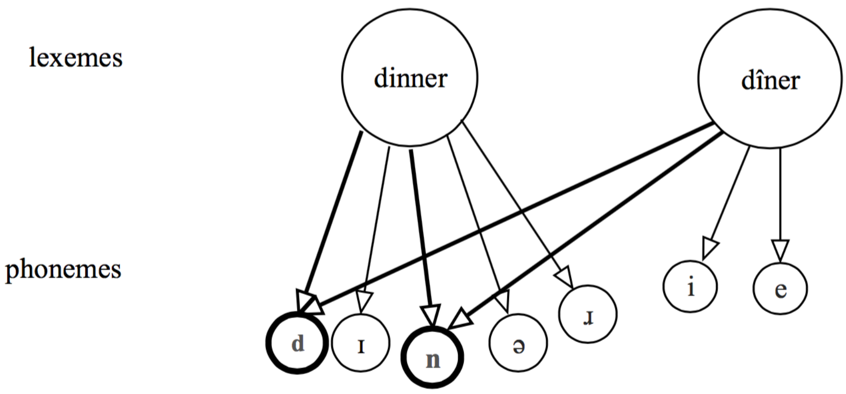


Figure 3.7: WEAVER++ for Bilingual Speakers

In addition to this, the non-native WEAVER++ model introduces control processes; that is, when speaking in an L2, the speaker places a goal symbol in working memory which specifies to speak in the L2, and this goal symbol is activated at every stage of production.[15]

While WEAVER++ displays competence in modeling both native and non-native production, little is said about the generation of prosodic frames in L2 speakers. One of the claims of WEAVER++ is that word forms activate a word's structural and segmental components separately, on the basis of contextual information for the utterance, and the structural component (the prosodic frame) is the Phonological Word frame. If we assume that WEAVER++ allows for the activation and spreading of non-target lexemes down into the word-form encoding processes, do we also assume that the default prosodic frames of the non-target language are activated?

Deviant syllabification, phrasing, and rhythm are all marked features of non-native language use: for example, Colloquial Bengali lacks consonant clusters on the edge of words, which sometimes results in errors such as [grun] for [graʊnd] in English.[16] Experimental evidence bolsters this argument: Peperkamp et al. (1999) found that "phonologically ungrammatical" sounds in an L2 were often assimilated into the closest acceptable form in the L1. This suggests firstly, that there is interference from non-selected L1 representations during phonological encoding, and secondly, that speakers exert some conscious control over this process. This raises the possibility that there is indeed interference from the non-target (L1) language's prosodic structure during L2 phonological encoding.

3.4 Summary

This chapter has traced language production modelling in Psycholinguistics from its earliest roots. We have seen the role that speech error data has played in the construction of models, and we have also seen how the models have been adapted and grown in

[15]Evidence for these control processes comes from a number of L1-L2 picture-word interference tasks, where interference effects appear at the intended stages (c.f. Hoshino & Thierry, 2011; Runnqvist & Costa, 2012; Shao et al., 2013; and Paivio, 2014).

[16]Lahiri, personal communication, 2015.

order to account for new experimental data over the years. The association of speech errors to discrete units of speech production was a crucial step in the development of production modelling: theorists were able to assign errors to specific processes during the generation of language. Production modelling began with very strict, modular models in which lexicalisation was the major focus and word-form encoding was seen as a simple process where grammatical and phonological information was assigned to words. Through the evolution of models, we have seen the presentation of spreading-activation, slot-and-filler, and interactive models that attempt to account for more numerous and complex processes in production. In particular, we have seen the introduction of models that view the phonological encoding process as a series of intermediate stages, as well as the identification of an underlying representation that undergoes phonological processing.

We have also observed how existing models deal with the grammatical and phonological encoding stages of production in non-native speakers. As with the native models, the most successful non-native models are those that are capable of explaining a variety of evidentiary data (speech errors and experimental tasks), coping with the features of connected, natural language, and illustrating in detail all levels of language production. WEAVER++ displays competence in all these areas without requiring the addition of numerous L2-specific processes. In addition to being able to account for encoding times in psycholinguistic data, it is able to handle features of unbalanced non-native speakers by specifying the amount of activation the L2 lemma and word-form receive.

In Chapter 2, we discussed the theory behind what is being planned: that is, how the units of spoken language are structured. Using the framework presented in Prosodic Phonology, a theory of hierarchically-grouped prosodic constituents, we analysed how speech was organised at the word-level. We focused especially on the Phonological Word constituent, which Nespor and Vogel (1986) described as the constituent where phonological, morphological and syntactic processes converge. As we have seen from models such as Levelt (1989, 1999) and Roelofs (1997), this constituent also has its place in the cognitive architecture of speech production: these models assume that a sub-stage of the phonological encoding process is dedicated to the arrangement of segments into specific frames. These frames are not lexical word frames, but rather Phonological Word frames. Now we have a foundation from which to move forward: we know what speakers are planning, at least in the post-lexical stage, and we know where this planning is taking place (at a specific substage of phonological encoding). This leads us to the experimental component of this study.

4

General Methodology

4.1 Overview

This study contains data from eight speech production experiments. The purpose of this work was to elicit reaction time evidence for the planning and production of prosodic units in native and non-native speakers of English. This was done by varying the number of Phonological Words in the target utterance in a prepared speech paradigm. The design of these tasks is based on those in Sternberg et al. (1978), Wheeldon and Lahiri (1997), and Wheeldon and Lahiri (2002). Because the experimental method used in this study remains constant throughout all eight experiments, in the interest of brevity it will be described in detail in this chapter.

4.2 Stimuli Selection

The main aim of this study was to examine the planning of post-lexical units in speech production: therefore, the conditions in each experiment were constructed from hypotheses about both the principles in prosodic constituent theory and models of speech production. In particular, we wanted to elicit data that reflected the phonological planning process. As discussed both in the Introduction and Chapter 2, the nonisomorphy of syntax and phonology at word level is visible in the structures themselves:

4.1 (a) **syntactic structure:** $[\text{I}]_{NP}$ $[[\text{took}]_{V}$ $[[\text{a}]_{Det}$ $[\text{bath}]]_{NP}]_{VP}$

(b) **prosodic structure:** $(\text{I})_{\omega}$ $(\text{took a})_{\omega}$ $(\text{bath})_{\omega}$

This asymmetry is particularly observable in prosodic structures containing compounds or clitics. It is for this reason that we chose to test these structures as our experimental items. Accompanying each experiment in this study are a set of hypotheses built on information we have gathered from previous reaction time experiments as well as theories about the shape of the Phonological Word in English. Each set of hypotheses addresses the experimental task type, the speaker group, and the stimuli themselves.

Many psycholinguistic tasks have exploited compounds as useful tools to examine morphological planning and structure in both the production and comprehension of language. The majority of these studies have focused on the lexicalisation of compounds: that is, how (or even if) they are stored in the mental lexicon (Griffen & Bock, 1998; Jescheniak & Levelt, 1994; Bates et al., 2003; and Bien et al., 2005). These tasks have focused heavily on the issues surrounding the word frequency of compounds. The effect of frequency on production latency in general has long since been established: simply put, words with higher lexical frequency are accessed faster than words with low frequency (cf. Oldfield and Wingfield, 1965). However, compound words present a complication: what affects production time for these items? Is it the individual morphemes of the compound?

The frequency of the compound as a single word? The sum of the morphemes? The first morpheme? The second?

A number of eye-tracking, lexical decision, and naming tasks have been used to study the effect of compound word frequency in an attempt to understand how compounds are stored and prepared during speech production (Pollatsek et al., 2000, De Jong et al., 2002, Juhasz et al., 2003, Duñabeitia et al., 2007). The findings of these studies have been somewhat ambiguous, particularly in regards to models of lexical encoding. We saw in Chapter 3 that spreading activation models (e.g. Dell, 1986; Roelofs, 1997) conceive of production as a series of levels in which nodes receive activation and spread to other semi-related nodes. Non-decompositional theories of lexical encoding (Collins & Loftus, 1975; Roelofs, 1992, 1993, 1997; Baayen et al., 1997; Mahon et al., 2007) assume that compound words are treated as a **single node** from conceptualisation to the initiation of phonological encoding. Therefore, only the simple compound frequency (that is, the frequency of the compound as a whole) should have an effect on production. In the strictest form of this theory, even the frequencies for the plural form of the compound (e.g. *doghouses*) are considered separate from the singular form (*doghouse*) (as in Figure 4.1):

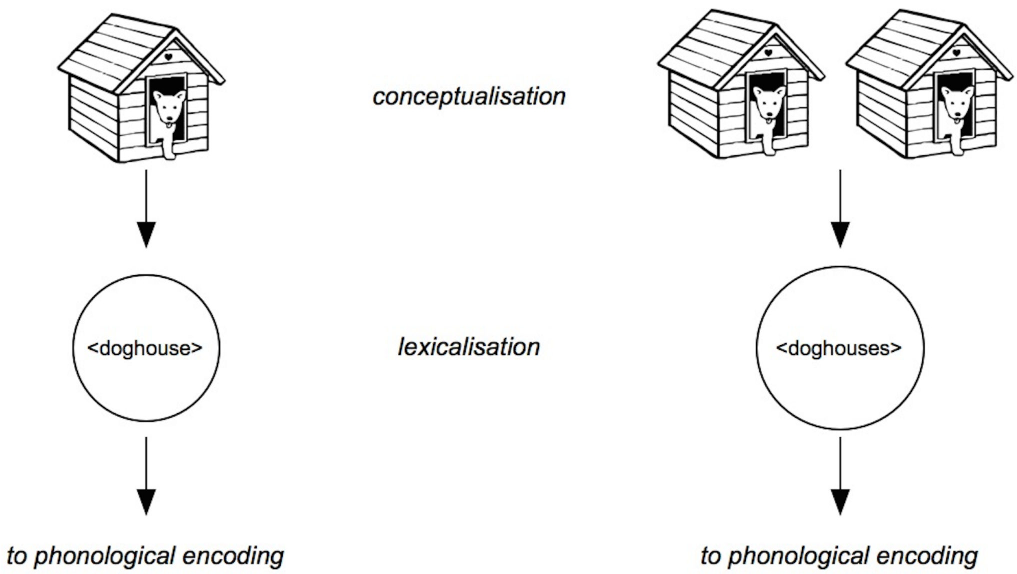


Figure 4.1: Planning of "doghouse" and "doghouses" in Non-Decompositional Models

Decompositional theories of lexical encoding, on the other hand, treat compounds as decomposed nodes at the lexeme level (Dell, 1986; Finkbeiner et al., 2006; Levelt et al., 1999; Levelt, 2001). In this case, production time should depend on the individual morphemes of the compound (*dog* and *house*) (see Figure 4.2).

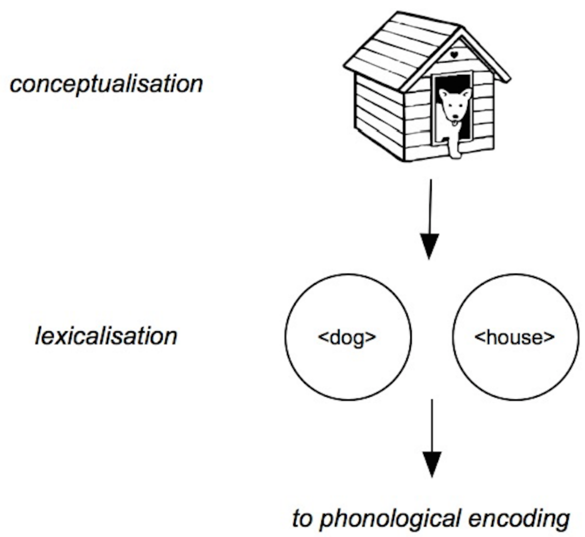


Figure 4.2: Planning of "doghouse" in Fully Decompositional Models

Experimental evidence regarding the storage of compounds varies widely. Hittmair-Delazer et al. (1994) and Semenza et al. (1997) observed that anomic patients often tended to replace noun-noun target compounds with other noun-noun compounds in naming tasks, while Delazer & Semenza (1998) also reported on an anomic patient who showed regular signs of separate lexicalisation of the compound's components. This patient could regularly retrieve one morpheme of the Italian compound with ease, but would then substitute the other with a component that was neither semantically nor phonologically related to the target's components: e.g. *bidonerifiuti* (lit. 'bin rubbish') for *portarifiuti* ('dustbin'). Juhasz et al.'s (2003) results suggested that only the compound's final morpheme frequency had an effect on reaction times: in their tasks, compounds which contained low-frequency final constituents elicited significantly longer naming latencies (550 ms) than those that contained high-frequency final constituents (511 ms).

In contrast, Bien et al. (2005) found that only the summed frequencies (morpheme1 + morpheme2) affected production latencies. In their production task, they varied the frequencies of both the first and second constituent (as well as the simple frequencies) in Dutch compounds and only found a frequency effect for the sum of the head and modifier constituents. Conversely, in a lexical decision task, Duñabeitia et al. (2007) observed that reaction time was only affected by the final constituents in Spanish and Basque compounds.

While constituency-effect studies have provide a wide variety of insights about the storage and lexicalisation of compound words during language production, their lack of consistency only serves to muddy the already murky waters of compound structure.[1] At the very least, we can glean one thing from these studies: the possibility of a compound frequency effect cannot be ignored in production experiments. When measuring reaction times for the compound stimuli, we must entertain the possibility that frequency is affecting the construction of these items. Because of the disparate results listed above, the only option is to test for an effect of every possible compound frequency in our experiments: simple, summed, initial (first constituent), and final (second constituent). For example, if the simple compound word frequency is having an effect on production latency, then we will see that *bathroom* elicits the fastest response times for the compound word stimuli (cf. Figure 4.3).

[1]The issue is further complicated by other factors that may affect compound naming latencies: Janssen et al. (2014) associated the wide-ranging results in frequency studies such as those discussed above with a lack of control of input representation by the researchers. For example, they claim that if the representation is displayed in an orthographic or phonological string that contains the compound's constituents, it causes them to be immediately recoverable from the lexicon.

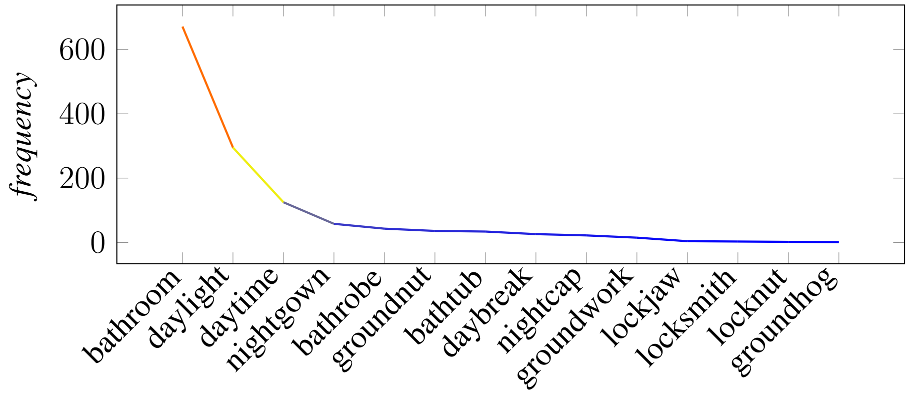


Figure 4.3: Compounds ordered by Simple Compound Frequency

If only the frequency of the first unit predicts mean latencies, then all compounds containing *day* will have the fastest latencies (cf. Figure 4.4).

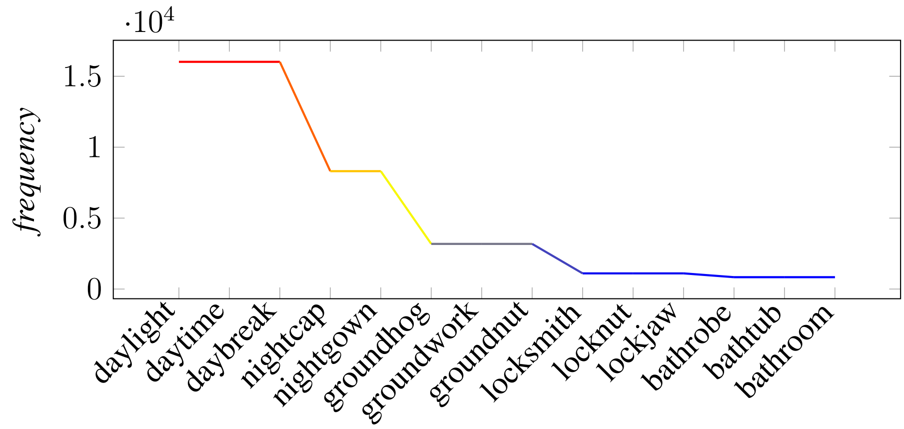


Figure 4.4: Compounds ordered by First Morpheme Frequency

If the naming latencies for the compounds differ significantly, this may indicate that frequency is an effect, and thus must be taken into consideration in the analysis.

4.3 Design

The measurement of reaction times has been a successful method of observing different cognitive processes involved in planning speech (Damian et al., 2001; Oldfield and Wingfield, 1965; Balota & Chumbley, 1985; Balota et al., 1989; Jescheniak &

Levelt, 1994; Bien et al., 2005; Alario et al., 2010), dual-task interference (Pashler, 1994; Ferreira & Pashler, 2002; Roelofs 2008), and morphological processing (Buck-Gengler et al., 2004). For example, we have seen evidence from word-distractor tasks that indicates interference from words that share grammatical class (Schriefers et al., 1991) or syntactic features (Schriefers, 1993). We also know that interference can occur if words share sounds or segments (Meyer & Schriefers, 1991 and Schiller & Costa, 2006). Evidence from speech errors and experimental studies show that a single concept can also erroneously activate two different (but semantically-related) lemmas (Schriefers et al., 1990 and Garrett, 1992).

In order to examine the way in which the prosodic frames of language are generated, we must utilise a method that elicits data about the stage in which these features are prepared: phonological encoding. The paradigm designed by Sternberg et al. (1978) and extended by Wheeldon & Lahiri (1997, 2002) has generated substantial results for phonological encoding in native Dutch speakers. This study adapts the paradigm to test two new areas in production: phonological encoding in both native and non-native speakers, and the planning of complex morphosyntactic structures in English. Experiments 1 and 2 were designed to evaluate the effects of prosodic structure on naming latencies for native speakers of English. Existing evidence for the effect of planning time on phonological encoding (Balota & Chumbley, 1985; Balota et al., 1989) suggests that it is important to consider all possible planning environments. Therefore, Experiment 1 presented the target material in delayed task conditions, while Experiment 2 removed the planning time to present the target material in online task conditions. The duality of task type was retained throughout the study, with all experiments being run in both delayed and online task conditions. Experiments 3 and 4 were designed to increase the prosodic complexity of the target items by adding clitics. Experiments 5, 6, 7 and 8 examined naming latencies in these items when planned by non-native speakers of English. Thus the full battery of experiments took the form shown in Table 4.1.

Table 4.1: Experimental Setup

Stimuli	Task	Native	Non-Native
Compounds	Delayed	Exp 1	Exp 5
	Online	Exp 2	Exp 6
Compounds + Clitics	Delayed	Exp 3	Exp 7
	Online	Exp 4	Exp 8

In each experiment, participants heard a question prompt and were required to respond using the target in an utterance. In Experiments 1, 2, 5, and 6, speakers were given a single, "simple" prompt:

Simple Prompt:
What was it?

This experimental question was recorded in a soundproofed booth by a male Southern British English speaker.

In the compound-clitic experiments, speakers heard one of five alternating "complex" prompts: their task was to answer that question in the prompt using the target word or phrase on the screen:

Complex Prompt:
What are clean?
What are good?
What are big?
What are great?
What are nice?

These experimental questions were recorded in a soundproofed booth by a female Southern British English speaker.

4.4 Subjects

4.4.1 Native Speakers

Native speakers were recruited through mailing lists and college advertisements at the University of Oxford. Native speakers were defined as individuals from the

United Kingdom who had spoken British English since birth and did not have fluency in any other languages.

4.4.2 Non-Native Speakers

For the non-native speaker group, we recruited native speakers of Standard Colloquial Bengali. The reasons for this are twofold. The first reason is the relationship between the two languages: compounding is very productive in both English and Bengali. However, the two languages also differ enough to be experimentally useful: for example, while noun-noun compounds are very common in Bengali, they display properties that English compounds cannot (e.g. inflection on the initial morpheme). This will be especially instrumental in the experiments testing cliticisation, as Bengali also exhibits different clitic attachment rules (cf. Fitzpatrick-Cole, 1990).

The second reason for choosing this speaker group was language background: all Bengali speakers in this study had the same nationality, the same education level and background, and were from the same age bracket. The importance of establishing language background for studies on non-native speakers is now generally recognised: without knowledge of the speakers' backgrounds, we have no concrete way to define fluency (cf Grosjean et al., 2004).[2] Therefore, a language background questionnaire based on the LHQ ("Language History Questionnaire", Li et al., 2014) was administered (see Appendix A). This questionnaire contained questions about age of acquisition (AoA), language use, educational background, and fluency.

To test proficiency levels in English, all participants were given a speaking and reading exercise based on the International English Language Testing System (IELTS) (British Council, 2013). The exercise measured coherence, lexical resource (vocabulary), grammatical range, accuracy and pronunciation. The researcher asked the participant to discuss simple "everyday" topics (such as holidays, friends, and family). It also

[2] "Ratings of proficiency alone are not sufficient to determine bilingual language status and that bilinguals' language learning and language use experiences play a significant role in shaping their linguistic competence" (Marian et al., 2007).

contained a reading component in which the participant was required to read a paragraph. Speakers were able to show a high level of oral proficiency with good pronunciation, a wide vocabulary, ease of speaking, and wide grammatical range. Examples of these tasks can be found in Appendix A.

Data was collected during two fieldwork trips to Kolkata, India in March, 2012 and February, 2015. Speakers ranged from age 14-28, were female, and were students from advanced learning institutions in Kolkata (Gokhale Memorial Girls' College and Shri Shikshayatan School). Selection for the non-native speakers was based on the participants' medium of instruction during primary and secondary school levels. In West Bengal, students follow one of two tracks: English-medium or Bengali-medium. English-medium students receive instruction in English from Class 1 (age 6), but speak Bengali or Hindi at home and with their peers, while Bengali-medium students receive all of their education in Bengali.[3] Experiment 5 contained an additional speaker group: 18 low-fluency English speakers. This group received instruction in Bengali at school and had generally low proficiency in English; it was used to test for any fluency effects in the non-native experiments.

4.5 Procedure

In all eight experiments, participants received information both auditorily and visually. The programme used to present the stimuli was E-Prime (2002), an experimental design programme run on a Pentium P6200 PC laptop. Experimental questions were always present auditorily, while the targets were presented on a single screen in lowercase, black 18-point Courier New font on a white screen. Participants were tested individually. Native speakers were seated in a recording booth in the Language and Brain Laboratory in Oxford, while non-native speakers were seated in a quiet, empty classroom in Kolkata.

[3] In the early 1980s, the Left Front government sought to abolish the teaching of English in primary schools in West Bengal. However due to pressure related to the globalisation of English and the rise of the IT industry, the government chose to reintroduce English in 1992 (Sen, 2015).

Each participant was fully briefed about the nature of the experiment both on paper and in person, and were given the option to leave the experiment at any time.

Before the experimental trials began, participants completed a block of practice trials: they were able to repeat these as many times as they wished. The target items in the practice block did not contain any of the experimental stimuli, but used similar words and phrases. During the practice trials, the experimenter remained in the room to assist the participant with any questions. Then the participant was left alone. At the end of each block of items, the participant was offered an unscripted break. The next block began when the participant pressed any key on the computer.

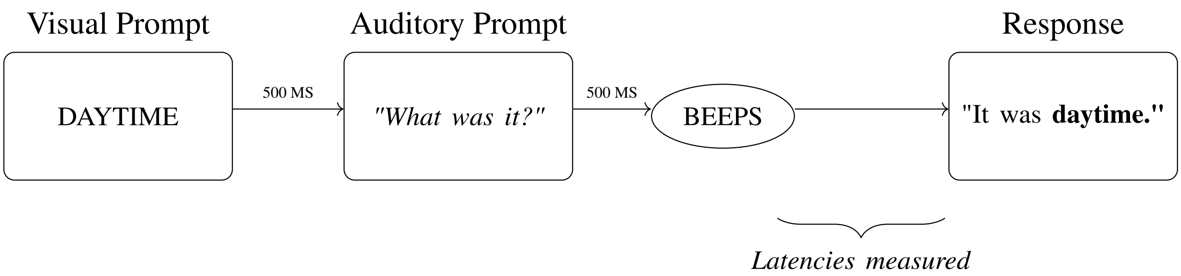


Figure 4.5: Delayed Task Design: Compounds

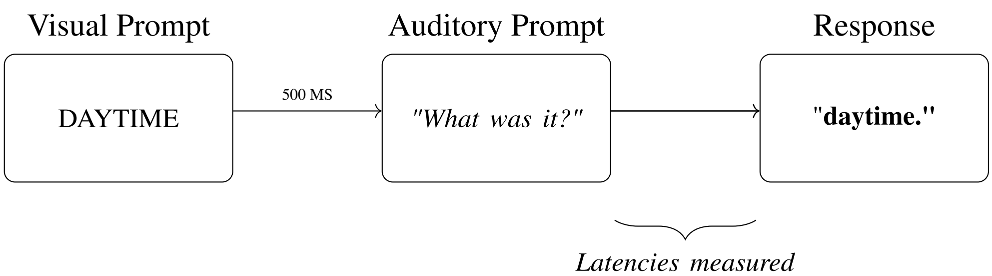


Figure 4.6: Online Task Design: Compounds

In the compound tasks (Figures 4.5 and 4.6), each trial began with a blank screen, shown for 500 ms. Then a fixation cross appeared centred on the screen for 500 ms. After the fixation cross disappeared, the target item was shown centred on the screen for 500 ms. The word then disappeared from the screen for 500 ms and then the participant heard the experimental question, *"What was it?"* through the headphones.

In the delayed task (Figure 4.1), this was followed by three beeps of equal duration: the first occurred 2 seconds after the offset of the word, the next 1 second later, and the final beep occurred at a variable latency (800 ms, 1200 ms, or 1400 ms) from the offset of the second beep. These variable latencies were chosen in order to prevent participants from guessing when the prompt would occur, and falling into a pattern. In the online task, the design remained the same, save for removal of preparation time: speakers were told to answer immediately after hearing the auditory prompt. The design for the compound + clitic tasks (Figures 4.7 and 4.8) was identical; only the auditory prompts and intended responses were changed.

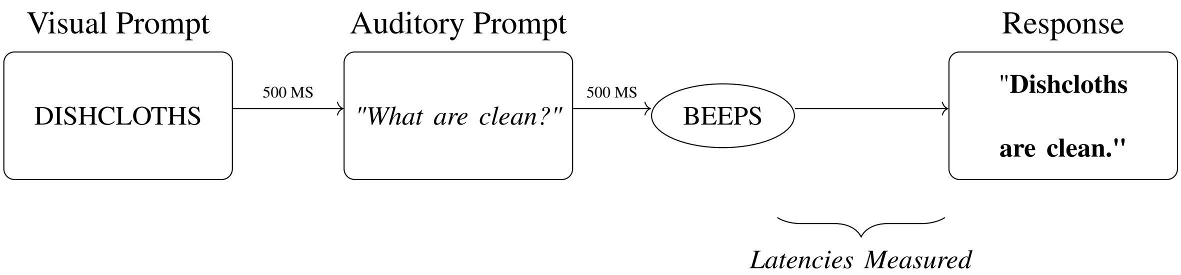


Figure 4.7: Delayed Task Design: Compounds + Clitics

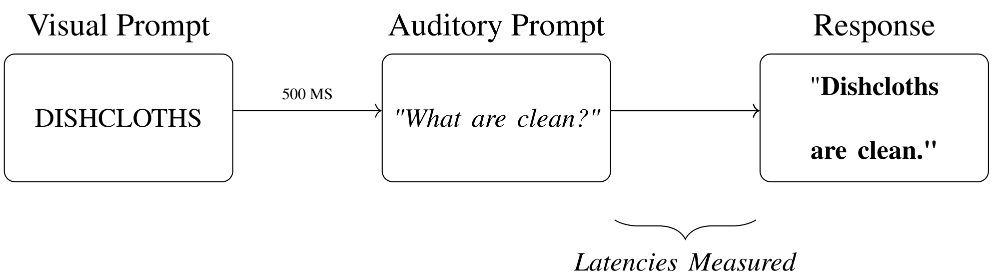


Figure 4.8: Online Task Design: Compounds + Clitics

The responses were recorded directly on a Sony DTC-100 ES DAT-recorder using a dual-channel headset that also contained a microphone. This allowed the microphone to stay at a relatively-constant distance from the mouth, and to minimise interference from background noise in the Kolkata locations. Recording continued throughout the experiment. The audio data was sampled at 44.1 kHz in two files which contained, respectively, the experimental output (phrase and beep) and the subjects' microphone input. These audio files were then converted to one audio file (with the experimental output on one channel, the speaker on the other channel) and an accompanying Praat (2013) TextGrid file of the target labels generated by a Perl-program designed by Henning Reetz (2010). As per Rastle & Davis (2002)[4], these labels were all checked visually and adjusted if incorrect (i.e. if the script identified a non-speech sound as the boundary). An example of this process can be seen in Figure 4.9 below.

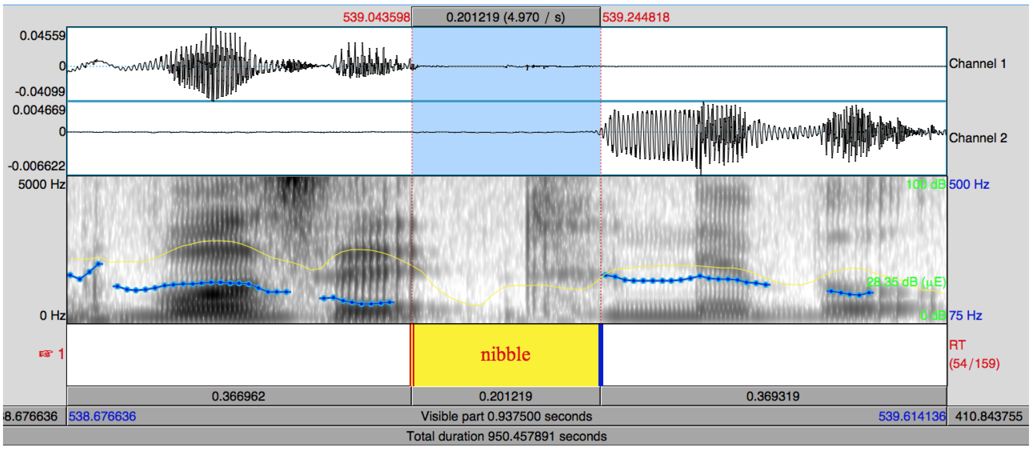


Figure 4.9: Example of a Measurement

Response latencies[5] were defined as the delay between either the third beep (in the delayed tasks), or the release of the final sound in the auditory prompt (in the online tasks). As we were only interested in the effect of the prosodic construction of the target words, durations of the utterances were not measured in these experiments.

[4] "We would also argue that more accurate measurements of acoustic onset can be derived by visual inspection of the acoustic waveform..."(313).

[5] The terms "response latencies", "naming latencies", and "reaction times (RT)" will be used interchangeably throughout this study.

4.6 Task Considerations

It is worthwhile to note that the experimental tasks in this study involve both visual word recognition and production: the participant sees a target word or phrase on the computer screen and must produce a response containing that target. In some tasks, this is simply the target itself (e.g. *groundhog*); in others, it is a complex utterance requiring both the target and the correct response to the prompt (*groundhogs are nice*). We have discussed the processes involved in production at length in Chapter 3; however, it seems appropriate to briefly discuss the relationship between speech recognition and production here.

Recognition has been described as the "reverse" of the production process; as such, similar issues arise in the literature for both processes (Bock and Griffen, 2000). Modern models of reading (visual word recognition) resemble the computational production models, approaching the data either with connectionist or interactive modelling processes. While production begins with the conceptualisation of a word, visual word recognition begins with orthographic identification and activation of word meaning (Frost, 2012). We know that phonological processing is involved in both reading and speaking: data from masked priming tasks has shown that the phonology of a word is accessed between 50-100 ms after words are presented (Ferrand & Grainger, 1993).[6] There is evidence that monolingual readers activate phonological representations while reading even if they are not planning spoken utterances (Perfetti & Tan, 1998; Rastle & Brysbaert, 2006). More recently, Wheat et al. (2010) found evidence of early phonological activation during visual word recognition in induced priming tasks, "consistent with phonological access being mediated by a speech production code" (5229).

The degree to which phonology is activated in the visual recognition process is

[6] Additional evidence has been found in relation to phonology and writing: Qu et al. (2015) found evidence for the involvement of phonology in handwritten word production during masked priming experiments. They tested native Mandarin speakers, having them write responses to target picture stimuli: notably, the Chinese orthographic system contains "compounds of multiple characters, yet most characters themselves are morphemes and hence carry meaning". The fact that a priming effect was elicited even when the conversion route between orthography and phonology was not 1 to 1 suggests, they propose, "a direct link between spoken syllables and written characters" (6).

uncertain. However, most approaches acknowledge at least some interaction between the two operations (Lukatela & Turvey, 2000; Pexman et al., 2001, 2002). Unlike spoken word recognition which must be able to identify words with numerous phonological variations (cf. Lahiri & Marslen-Wilson, 1991), orthographic recognition involves a more-direct activation of semantic information. Orthographic targets contain clear word boundaries and all of their elements are presented simultaneously; as Hannagan et al. (2013) write, "visual words have several advantages over spoken words as objects of perception". As for non-native word recognition, research has shown evidence of the activation of both the L1 and L2 languages (de Groot & Nas, 1991; Grainger & Frenck-Mestre, 1998). L2 processing models (e.g. Dijkstra et al. 2000, 2002) assume spreading activation of the non-target (L1) language throughout the recognition operations. This study does not attempt to elicit evidence for any particular model of visual word recognition in native or non-native English speakers; it will, however, take into account processing demands for both task type and fluency level in the different groups of speakers.

4.7 Analysis

Each analysis section presents the statistical analysis of reaction time data, mean naming latencies by condition and (if applicable) preparation deadline, a comparison of density estimates for each condition, error patterns, error rates (by condition and, if applicable, preparation time), frequency analyses, and word length analyses.

Results were cleaned using the JMP (11.0) Statistical Package and SPSS statistical software. All data points beyond two standard deviations from the mean were considered outliers and removed. Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were also discarded. Responses for which the participant responded before the deadline (i.e. the beep in the delayed task, or before the question ended in the online task) were removed as well. One stimulus (the compound *nightmare*) was removed from Experiments 1 and 2 because it did not comply with the compound criteria of transparency. Participants that returned an exceptionally-high number of errors or response times outside the group mean of each experiment were excluded as well.

4.7.1 Naming Latencies

We used R Statistical Software (R Core Team, 2015) and the 'lme4' package (Bates et al., ress, 2015) to perform a linear mixed effects analysis of the relationship between reaction time (naming latency) and condition type. Linear regression models, and linear mixed effects models in particular, are especially attractive for use with linguistic data. Designed to model groups of data as well as different response variables, they work particularly well for the by-subject and by-item analyses we present in this study. Rather than being forced to run separate ANOVAs for subject and item analyses, the mixed effect models are able to handle both of these variables as random factors in a single analysis. In the case of the experimental tasks presented in this study, we are able to analyse the effect of a number of variables (e.g. condition type and preparation time) on RT in one model. Furthermore, linear mixed effects models can account for variation in the intercepts and/or slopes for the effect of condition on both the subjects and the items presented in the task.

As fixed effects, we entered reaction time, condition type, and (in the delayed tasks) preparation time. As random effects, we had subjects and items. Each analysis will present the results of the maximally-converging model (as per Bates et al., 2015), as well as a discussion of those models that did not converge. Following Baayen (2008), t-values and standard errors are reported for significance of effect. There are no p-values reported by the 'lme4' package; this is due to the fact that the fixed-effect parameters of mixed effect models are not as easily calculated as those for models containing only fixed effects.[7] Further, in experiments containing large data sets, Baayen states that significance can be assumed for any absolute t-statistic exceeding 2:

[7] "It is not even obvious how to count the number of parameters in a mixed-effects model" (Baayen, 2008). See also Douglas Bates' 2006 post to the [R] forum, in which he baulks at the "obviously correct approach of attempting to reproduce the results provided by SAS" by providing p-values.

> "For data sets characteristic for studies of memory and language, which typically comprise many hundreds or thousands of observations, the particular value of the number of degrees of freedom is not much of an issue... For such large numbers of degrees of freedom, the t distribution has converged, for all practical purposes, to the standard normal distribution. For large data sets, significance at the 5% level in a two-tailed test for the fixed effects coefficients can therefore be gauged informally by checking the summary for whether the absolute value of the t-statistic exceeds 2" (2008: 398).

For the reasons stated above, we feel comfortable in only reporting t-statistics for our experiments, and assigning significance to any value above 2.

4.7.2 Diagnostics

Following Winter (2013), a number of diagnostics were carried out for each analysis to check goodness-of-fit of the model to the data: these included tests for normality of residuals, absence of collinearity, and independence. For normality of residuals, scatterplots, histograms, and quantile-quantile (QQ) plots were visually inspected with respect to the presence or absence of patterns of heteroscedasticity (that is, when variables are not equally-variable across a range of values of another variable that predicts it). Heteroscedasticity can signify that a model is fitted poorly to the data. Examples of homoscedasticity and heteroscedasticity are given, respectively, in Figures 4.10 and 4.11 below.[8]

[8]Examples in Figure 4.10 were generated from the "LakeHuron" dataset of the "datasets" R package (R Core Team, 2015), while examples in 4.11 were generated from random data points.

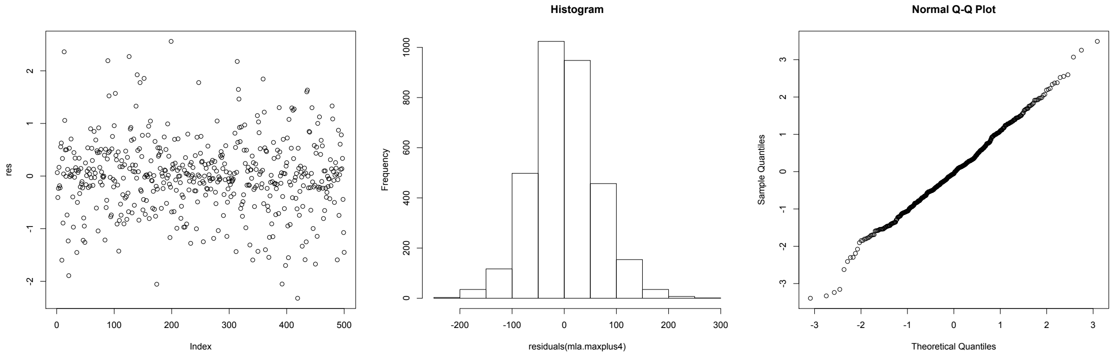


Figure 4.10: Examples of Homoscedasticity in Residual Plots

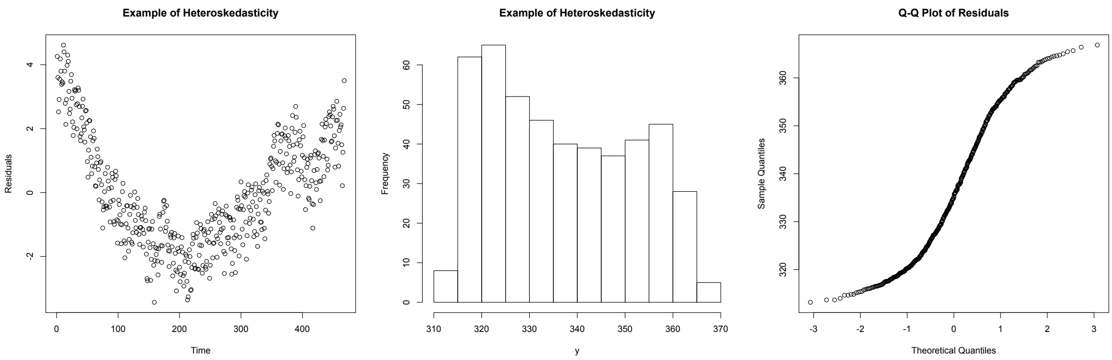


Figure 4.11: Examples of Heteroscedasticity in Residual Plots

Figure 4.11 presents examples of heteroscedasticity in three different plots: scatterplots, histograms, and QQ-plots. One obvious sign of a heteroscedastic distribution is "coning" (in which the spread of data points widens as the value of the independent variable increases) (Taylor, 2014). Another sign can be seen in the scatterplot in Figure 4.11, where the data points follow a clear, u-shaped pattern: this signifies that the data is not in a standard normal distribution. Other indications of a poor fit include the distribution of data in a non-bell shape in a histogram and a curved line in the quantile-quantile (QQ) plot.

If the diagnostics discussed above identified a heteroscedastic distribution of data in our study, steps were taken to ratify this: following Winter (2013), we performed a log-transformation of the data to see if it improved the fit. Profile log-likelihoods of reaction times were generated via a Box-Cox power transformation from the "Car" package (Fox et al., 2009). If this resulted in a better distribution of data (as was the case in Experiments 5 and 6), the log-transform of the reaction times was used in the models. In analyses where log-transform values are used, this will be discussed thoroughly.

4.7.3 Errors

Error rates were reported for each condition, as well as for each preparation deadline. For all experiments a normal generalised linear model (GLM) was used, in which reaction time was treated as the fixed effect. We distinguished five types of errors: disfluency, null response, stress error, voice key malfunction, and wrong item. Disfluencies included stuttering, repetition, false starts, fillers, or slips of the tongue. Stress errors involved the incorrect placement of either initial or final stress on any segment of the stimulus. Voice key errors included popping, errors in recording, or any other equipment failure. Wrong items were classified as well-formed words that were simply incorrect (e.g. mistaking *bathroom* for *bathtub*). These classifications became vital for our analyses of non-native speech.

4.7.4 Frequencies and Word Length

In the frequency analyses, the RT data were again submitted to a linear mixed models analysis, in which frequency, condition, and preparation time (beep) were all treated as the fixed effect factors. Each possible compound frequency was tested: simple (CELEX and log), left morpheme, right morpheme, and summed frequency. In the case where a frequency effect was found, this was included in the initial reaction time model described in §4.7.1.

Word length (number of letters) has also been found to affect naming latencies for both word recognition and naming (cf. James, 1975; Whaley, 1978). Therefore, we also looked for an interaction between word length and condition type on reaction times.

Again, as with frequency, if an effect was found, this was incorporated into the RT model.

4.8 Summary

This chapter has served to introduce the experimental method, design, and analysis used throughout this study. Using the paradigm introduced by Sternberg et al. (1978) and further developed by Wheeldon & Lahiri (1997, 2002), we present eight experiments testing the preparation and production of complex morphosyntactic words in a variety of experimental task conditions (delayed and online) and speaker groups (native and non-native English speakers). Due to the effect of frequency on language production, the target stimuli were selected carefully with regards to their frequencies. Further, all possible frequencies for compound words will be considered in the analyses. In this chapter, we have also discussed the advantages of the statistical methodology for analysing these experiments: per Baayen (2008) and Bates et al. (2015), we see that linear mixed effect models are preferable models for analysing linguistic data, and production data in particular.[9] The following chapters aim to combine theoretical approaches with an established methodology in order to present an integrated study of the production of complex morphosyntactic structures in native and non-native speakers of English.

[9] "Essentially, all models are wrong, but some are useful; the practical question is how wrong do they have to be to not be useful" (George Box, 1987)

5

The Phonological Encoding of Compounds by Native English Speakers

5.1 Introduction

The experiments presented in this chapter seek to test two claims: that the unit of processing at the phonological encoding stage is a prosodic (not lexical) unit, and that compounds constitute a single, recursive prosodic unit in English. These claims are based on models of the phonological encoding process (Levelt, 1989; et al., 1999), as well as the framework of Prosodic Constituent Theory. Our focus in this chapter is on the preparation of complex morphosyntactic words in English, and to what extent the processing of a lexical word differs from that of a Phonological Word in these structures. We have seen evidence from Wheeldon and Lahiri (2002) that Dutch speakers plan noun-noun compounds as single prosodic units. Using the same experimental paradigm, we look to replicate the effect in English.

First, Experiment 1 (§5.2) will test the delayed planning of compounds in native British English speakers, in order to see if they are processing compounds as single prosodic units when they have time to plan their speech. A discussion of the results from this task can be found in §5.2.6. Experiment 2 (§5.3) is an online version of Experiment

1: it tests whether a loss of planning time affects the size or shape of the Phonological Word unit in English. Results and the accompanying discussion are presented in §5.3.7, followed by a general discussion about the implications of both experiments.

5.2 Experiment 1: Delayed Production of Compounds

The first experiment in this study consists of a delayed production task with a simple question-answer paradigm. The studies by Wheeldon & Lahiri (1997, 2002) in particular, which expanded upon Sternberg et al. (1978), have formed a significant foundation for the development of the hypotheses, design, and predictions of the experiments presented in this chapter.

5.2.1 Hypotheses

Based on the theories and experimental evidence presented thus far, the main hypotheses tested in this study are as follows:

1. The unit of processing at the phonological encoding stage in English is prosodic, not lexical.

Numerous psycholinguistic studies have shown that the post-lexical encoding (also called "phonological encoding") stage is responsible for infusing abstract lexical representations with phonological properties such as segmental ordering (Meyer (1990, 1991); Wheeldon and Levelt (1995); Wheeldon and Morgan (2002), syllabification (Ferrand et al. (1996, 1997); Schiller (1998, 2000); Schiller et al. (2002)), and prosodification (Ferreira (1993); Wheeldon and Lahiri (1997); Roelofs and Meyer (1998); Wheeldon and Lahiri (2002); Jescheniak et al. (2003); Cholin et al. (2004); Damian and Dumay (2007)). Only recently have language production models begun to address the shape of the underlying representation in post-lexical processing stages. Levelt's original model (1989) described the output of grammatical encoding as a tree structure (called "the surface structure") in which lemmas were first retrieved from the mental lexicon as sublexical and subsyllabic units, then positioned in independent structures (such as word and syllable skeletons) during phonological encoding. The output representation therefore contained templates for syllable constituents and Phonological Word segments.

In Levelt et al.'s WEAVER++ model (1999), each level accepts a prepared representation as input and generates a new representation as its output. During phonological encoding, prosodic structures are built by attaching segmental specifications to metrical frames during syllabification. Information about the utterance's phonology is combined with existing information in the morphosyntactic representation (as in Figure 5.1).

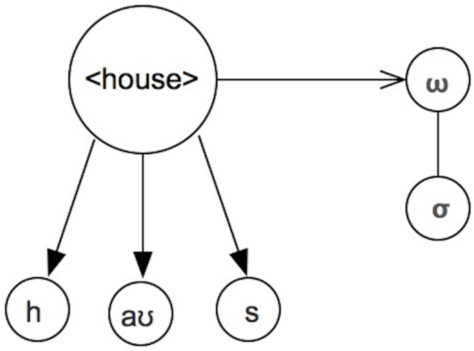


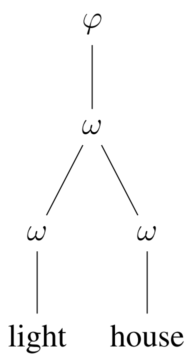
Figure 5.1: Phonological Encoding in WEAVER++

These models maintain that the final stages of language production are driven by prosodic, not lexical, units: that is, prosodic frames are arranged according to the prosodic structure of the utterance. As we saw in Chapter 2, a lexical word is a well-formed semantic and syntactic unit that can stand on its own, be uttered in isolation, and even be considered a full utterance. Conversely, a Phonological Word (PWd) is a prosodic unit that can contain a lexical word **plus** any number of unstressed items (e.g. Selkirk,1980b; Nespor & Vogel, 1986; Peperkamp, 1997). The number of lexical words is thus often different to the number of Phonological Words in an utterance. In this experiment, we predict that **only** the number of Phonological Word units (not the number of lexical words and/or syllables) in the target will predict naming latencies.

2. A compound word will be treated as a single, recursive Phonological Word for the purpose of planning speech.

In order to test Hypothesis 1, we need a target item that has a different number of lexical words than Phonological Words. Theories of phonological phrasing (e.g. Selkirk (1980a,b, 1986); Lahiri and Plank (2010)) assume that prosodic units are not isomorphic with syntactic units. For example, when treated as individual morphosyntactic units, *black* and *board* are separate phonological units with their own stress assignment (depending on the context in which they occur). When compounded, however, the units become one Phonological Word (*blackboard*) with a main stress falling on the first unit and a secondary stress on the second. Following Selkirk (1999) and Wheeldon and Lahiri (2002), we maintain that English compound words are recursive prosodic constituents, as in *lighthouse* in Example 5.1 below:

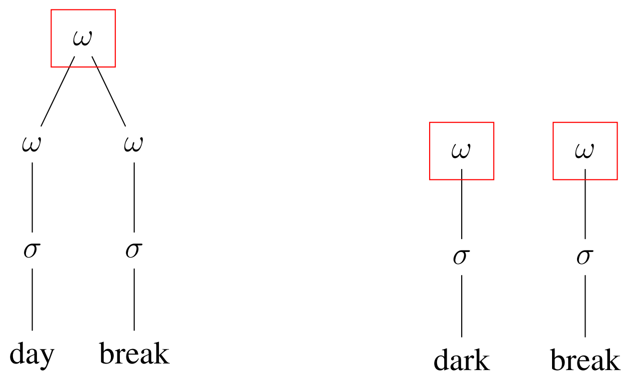
(5.1) the compound word *lighthouse*



Experimental evidence in support of the treatment of compound words as single prosodic units can be found in a number of prepared speech tasks (Wheeldon & Lahiri, 1997; 2002), picture-word experiments (Janssen, 2008), and implicit priming tasks (Jacobs and Dell (2014)). In Wheeldon & Lahiri (2002), noun-noun compounds (e.g *ooglid*, "eyelid') elicited significantly shorter results than adjective-noun phrases (e.g. *oud lid*, "old member"), even though they shared the same number of lexical words. Furthermore, the latencies for noun-noun compounds were similar to monomorphemic words (e.g. *orgel*, "organ"). More recently, Jacobs and Dell (2014) conducted a series of implicit-priming tasks that tested the production of simple words, phrases, and nominal compounds; their findings also suggested that compounds had priming behaviour similar to the monomorphemic words. That is, the onsets of the second unit of nominal compounds did not receive priming (e.g. the [d] in *sawdust*) while the same word did when used independently in a phrase (e.g. *grey dust*).

Following Wheeldon and Lahiri (2002), we present four conditions in which the number of Phonological Words differ from the number of lexical words: compounds (e.g. *bathroom*), phrases (*bad room*), simple words with initial stress (*basil*), and simple words with final stress (*balloon*). Following the evidence above, we predict that compounds will generate similar naming latencies to the morphosyntactically-simple words due to the fact that the unit predicting the naming latencies for compounds will not be the individual Phonological Words at the lower level of the structure, but the higher, recursive Phonological Word constituent seen in the tree structure for *daybreak*, as illustrated in example 5.2 below. Accordingly, phrases will elicit significantly different naming latencies to phrases, on account of their prosodic structure.

(5.2) Units predicting naming latencies in compounds and phrases



If these hypotheses prove correct, then they will lend evidence to the claim that the Phonological Word is the processing unit in phonological encoding, and will serve as a starting point for this study as a whole.

5.2.2 Stimulus Selection

The experimental materials were constructed from 60 items, 15 per condition type. Four experimental conditions were distinguished: noun-noun compounds, adjective-noun phrases, disyllabic morphosyntactically-simple words stressed on the first syllable, and disyllabic morphosyntactically-simple words stressed on the final syllable (see Table 5.1 for stimuli). Targets were arranged into five sets beginning with one of five phonemes: /d/, /g/, /n/, /l/, and /b/. These phonemes were chosen in order to more easily identify boundaries during analysis (cf. Rastle & Davis, 2002). 9 disyllabic filler words were included as well.

Table 5.1: Stimuli: Experiment 1

Condition 1: Compounds	**Condition 2: Phrases**	**Condition 3: Initial Stress**	**Condition 4: Final Stress**
daybreak	dark break	denim	default
daytime	dark time	debit	deceit
daylight	dark light	dagger	decree
groundhog	green hog	gravel	gazelle
groundwork	green work	griddle	gazette
groundnut	green nut	gravy	guitar
nightcap	nice cap	nibble	neglect
nightgown	nice gown	nickel	noblesse
nightmare	nice mare	nitrate	notate
lockjaw	long jaw	locker	lament
locksmith	long smith	locust	latrine
locknut	long nut	lodger	lapel
bathrobe	bad robe	banter	baboon
bathroom	bad room	ballad	balloon
bathtub	bad tub	basil	bamboo

The selection of target items was based on a number of considerations. First, because the overall aim of this study was to test phonological encoding in both native and non-native English speakers, the target stimuli needed to be made up of familiar words. Therefore, a judgement task was run on 35 native Southern British English speakers (an example can be found in Appendix B). This task collected rating data on all 60 experimental items as well as 12 filler items. Participants were asked to rate each word on a scale modelled on the Likert-Type Scale Response Anchors (Vagias, 2006). Participants rated their familiarity on a scale from 1 to 5, where 1 represented "not at all familiar" and 5 represented "extremely familiar". Results for the judgement task can be found in Table 5.2. Word familiarity was generally high for most items, with the exception of *groundnut*, *locknut*, and *noblesse*. Crucially, the mean judgement ratings for each condition did not significantly differ from each other (p=.0534).

As discussed in the General Methodology (Chapter 4), there are a number of other factors that can affect naming latencies in language production tasks. Therefore, measures of word frequency and word length were collected from the N-Watch statistical software (Davis. 2005). Frequency counts consisted of each item's CELEX (COBUILD) frequency, which included the combined written and spoken counts. Further measures included the (base 10 plus 1) Log Frequency (also CELEX) and word length (number of letters). While we have discussed the effect of frequency on language production, word length has also been found to affect both word recognition and naming (cf. James, 1975; Whaley, 1978).

The individual morphemes of compounds and phrases were matched as closely as possible for frequency and length: e.g. the first morpheme of the compound (e.g. *ground*) was matched for length and frequency to the adjective of the phrasal condition (*green*). To further account for any effects related to these factors, the second morpheme of the compound condition remained identical to the second morpheme of the phrasal condition (e.g. *groundHOG* and *green HOG*). Lastly, within the compound condition, the initial compound morpheme was identical for each phoneme-organised set: e.g. (*DAYbreak*, *DAYtime*, *DAYlight*). This was done in order to minimise any frequency effects between the compounds and phrases in the analysis. The adjective-noun phrases were also constructed in such a way to discourage speakers from compounding the two words in the phrase: therefore, they contained some unique combinations, such as *long jaw* and *green hog*.

The stress of the disyllabic initially-stressed words and disyllabic final-stressed words was confirmed by 35 native Southern British English participants. A questionnaire (examples of which can also be found in Appendix B) presented the experimental items along with filler words. Participants were asked to select which syllable of the word received the most stress, with the options "first", "second", and "not sure". Results for this task can be found in Table 5.4. Speaker judgements for the target items were generally strong, with all items rating at least 80% for correct stress.

Table 5.2: Weighted Averages of Word Familiarity: Exp 1

Cond 1:	Score	**Cond 2:**	Score	**Cond 3:**	Score	**Cond 4:**	Score
daybreak	5	dark break	5	denim	4.97	default	5
daytime	5	dark time	5	debit	4.84	deceit	4.91
daylight	5	dark light	5	dagger	5	decree	4.94
groundhog	4.84	green hog	5	gravel	5	gazelle	4.97
groundwork	4.88	green work	5	griddle	4.91	gazette	4.50
groundnut	2.88	green nut	5	gravy	5	guitar	5
nightcap	4.91	nice cap	5	nibble	4.84	neglect	5
nightgown	5	nice gown	5	nickel	4.84	noblesse	3.13
nightmare	5	nice mare	4.91	nitrate	4.31	notate	4.63
lockjaw	4.81	long jaw	5	locker	5	lament	4.84
locksmith	4.97	long smith	4.78	locust	4.91	latrine	4.78
locknut	2.84	long nut	5	lodger	4.44	lapel	4.94
bathrobe	5	bad robe	5	banter	4.94	baboon	4.97
bathroom	5	bad room	5	ballad	4.97	balloon	5
bathtub	5	bad tub	5	basil	5	bamboo	5
Mean Score:	4.675		4.979		4.866		4.774

Table 5.3: Mean Variables: Experiment 1

	Comp 1	**Comp 2**	**Comp**	**Adj**	**N**	**Sim Ini**	**Sim Fin**
CELEX freq	291.5	258.0	5.3	298.7	259.0	2.9	3.2
Log Freq	2.1	1.6	0.5	2.4	1.6	0.5	0.5
No of Letters	4.4	3.8	8.2	4	3.8	5.9	6.5

Table 5.4: Native Speaker Stress Judgements: Experiment 1

Item	**% Initial**	**% Final**	**% Unsure**	**Item**	**%Initial**	**%Final**	**% Unsure**
denim	80.00%	11.11 %	8.89 %	*default**	6.67 %	91.11 %	2.22 %
debit	81.82 %	9.09 %	9.09 %	*deceit*	2.27 %	93.18 %	4.55 %
dagger	88.64 %	6.82 %	4.55 %	*decree**	4.55 %	93.18 %	2.27 %
gravel	91.11 %	6.67 %	2.22 %	*gazelle*	13.64 %	84.09 %	2.27 %
griddle	82.22 %	2.22 %	15.56 %	*gazette*	6.67 %	93.33 %	0.00 %
gravy	88.89 %	2.22 %	8.89 %	*guitar*	2.22 %	91.11 %	6.67 %
nibble	82.22 %	11.11 %	6.67 %	*neglect**	11.36 %	88.64 %	0.00 %
nickel	86.67 %	6.67 %	6.67 %	*noblesse*	11.11 %	82.22 %	6.67 %
nitrate	82.22 %	8.89 %	8.89 %	*notate**	6.67 %	88.89 %	4.44 %
locker	86.67 %	6.67 %	6.67 %	*lament**	6.82 %	86.36 %	6.82 %
locust	86.67 %	6.67 %	6.67 %	*latrine*	8.89 %	86.67 %	4.44 %
lodger	84.44 %	8.89 %	6.67 %	*lapel*	9.09 %	86.33 %	4.55 %
banter	86.67 %	6.67 %	6.67 %	*baboon*	6.67 %	88.89 %	4.44 %
ballad	93.33 %	4.44 %	2.22 %	*balloon*	8.89 %	82.22 %	8.89 %
basil	91.11 %	6.67%	2.22 %	*bamboo*	17.78 %	82.22 %	0.00 %

5.2.3 Design

The experiment was designed using E-Prime 2.0 software. There were 9 blocks of 19 trials in total (171 items total including filler words) with optional breaks between the blocks. Items were distributed pseudo-randomly within the blocks, and three different versions of the experiment were presented. Each experimental word was presented once at each preparation latency: 800 ms, 1200 ms, and 1400 ms. Thus, each word appeared three times in the course of the experiment.

5.2.4 Participants

18 participants took part in Experiment 1. Participants were native speakers of British English and were recruited through mailing lists and college advertisements at the University of Oxford. Native speakers were defined as individuals from the United Kingdom who had spoken British English since birth; no participant reported having a second language. All speakers were between the ages of 18 and 26, and did not report any history of hearing or language problems. They were compensated for their participation. The experiment took place in a soundproof booth at the University of Oxford.

5.2.5 Results

The statistical analysis of the data collected in this experiment was carried out in the R statistical package (R Core Team, 2015). The reaction time data and frequencies were submitted to a linear mixed effects (LME) model analysis. §5.2.5.1 discusses the naming latency data, §5.2.5.2 discusses the error data, and §5.2.5.3 discusses frequency effects.

5.2.5.1 Latencies Analysis

Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error. Responses uttered before the final beep were also discarded. The total number of errors was low, with only 5.9% of data being discarded due to these parameters. Data trimming treated all data points beyond two standard deviations from the mean as outliers: this resulted in the removal of an additional 102 data points. The total loss made up 9.1% of data.

It was observed[1] that the target word "nightmare" is not fully transparent in meaning: only the word form *night* may be linked to its corresponding conceptual representation. Libben et al.(1997) argued that constituent priming effects for these items may be attributed to different sources: such "opaque" compounds possibly have different underlying representations at the conceptual level of production. There is therefore a possibility that speakers in this task will treat the target *nightmare* differently from the other, more-transparent compounds. For this reason, the target *nightmare* was removed from analysis.

The RT data were submitted to a linear mixed models analysis, in which naming latency (RT) was modelled as a function of the fixed effect factors, condition type (cond) and preparation time (beep). Subjects and items were treated as random factors. Following Bates et al. (2015), we began by testing the maximally-appropriate random structure and continued to adjust the model until it reached convergence. The maximal model contained a full interaction between condition and preparation time in the fixed effects, random slopes and intercepts for the interaction between condition and beep in the by-subject analysis, and random intercepts for condition and beep in a by-item analysis. This model[2] failed to converge. We then treated condition and preparation as separate fixed effects, but left the interaction in the by-subject analysis. This model[3] converged, but generated some problematic results: namely, the variance for preparation time was very small (1.701). This suggests that preparation time was having little effect on the by-subject or by-item analyses. The next maximal model to converge contained an interaction between condition and beep in the fixed effect factors, random slopes and intercepts for condition

[1] My gratitude to my transfer and confirmation examiners, who brought this to my attention.
[2] lmer(rt ~ cond*beep + (1+cond*beep|sub)+ (cond+beep|item)).
[3] lmer(rt ~ cond+beep + (1+cond*beep|sub)+ (cond+beep|item)).

in the by-subject analysis, and random intercepts in the by-item analysis.[4]

This model converged and its output is shown in Table 5.5, where significant effects are marked with an asterisk. This model appeared comfortably homoscedastic when fitted to residual plots, as shown in figure 5.6. There was some evidence of clustering in the scatterplot (Plot 1 in Figure 5.6); however, the histogram and QQ-plots were homoscedastic. Furthermore, a log-transform did not improve model fit. Following Baayen (2008), all t-values greater than 2 were treated as significant. The results from the mixed-effect analysis of condition on reaction time revealed a significant difference in the naming latencies for adjective-noun phrases (t=3.797*). In order to generate pairwise comparisons between the four conditions, each condition type was systematically treated as the intercept.

Adjective-noun phrases differed significantly from the other three conditions, while noun-noun compounds showed no significant difference to either of the morphologically-simple word conditions. These analyses also showed that simple words differed significantly from phrases, but not from one another. Mean reaction times and percentage error rates for Experiment 1 are shown in Table 5.6. The graph in Figure 5.7 shows the distribution of mean naming latency (RT) for all four conditions. Preparation time (beep) appeared to produce a significant effect on reaction times in an interaction with condition, but the effect did not reach significance in a pairwise comparison (all ts < 1.4). An analysis of the interaction of block and condition revealed no effect (t= 0.263), nor did a simple analysis of block on reaction time.

[4] lmer(rt ∼ cond*beep + (1+cond|sub)+ (1|item))

Table 5.5: Linear Mixed-Effects Analyses of RT Data in Experiment 1

	Estimate	**SE**	**t Score**
All conditions:			
(Intercept)	452.378	24.812	18.232
Adj-Noun Phrases	40.672	10.712	3.797*
Simple Initial Stress	-1.623	6.266	-0.259
Simple Final Stress	5.835	5.449	1.071
beep2	-11.752	4.279	-2.746*
beep3	-10.979	4.282	-2.564*
Pairwise Comparisons:			
Phrases as intercept:			
(Intercept)	493.050	29.832	16.528
Compounds	-40.672	10.712	-3.797*
Simple Initial	-42.295	14.212	-2.976*
Simple Final	-34.837	9.117	-3.821*
Simple initial as intercept:			
(Intercept)	450.755	23.673	19.041
Compounds	1.623	6.266	0.259
Phrases	42.295	14.212	2.976*
Simple Final	7.458	7.670	0.972
Simple final as intercept:			
(Intercept)	458.213	26.687	17.170
Compounds	-5.835	5.449	-1.071
Phrases	34.837	9.117	3.821*
Simple Initial	-7.458	7.670	-0.972

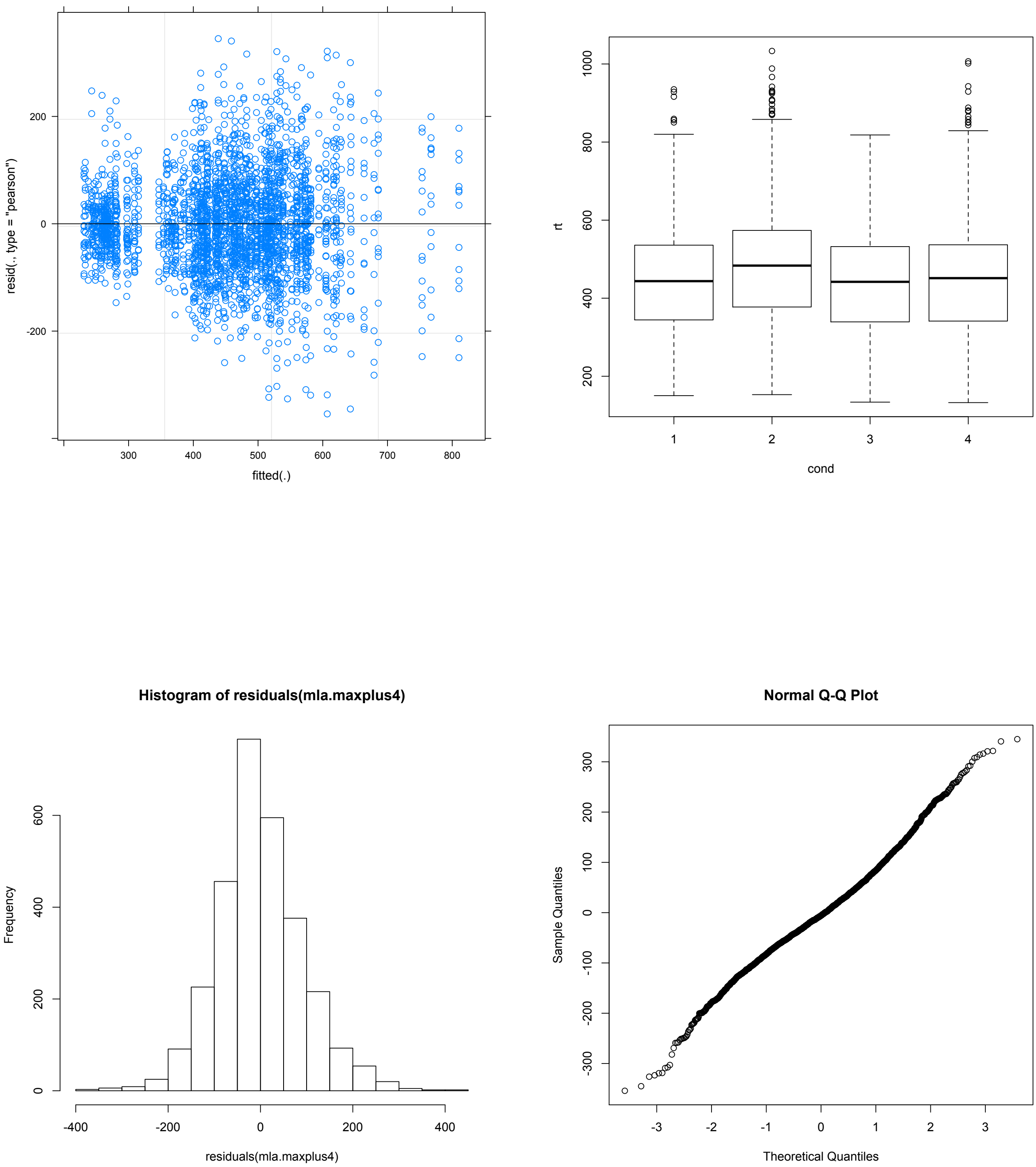


Figure 5.2: Residuals Analyses: Experiment 1

Table 5.6: Mean Naming Latencies (in ms): Experiment 1

	(1) Compounds	**(2) Phrases**	**(3) Initial Stress**	**(4) Final Stress**	**Mean Lat. (ms)**
PWds	1	2	1	1	
LexWds	2	2	1	1	
Syllables	2	2	2	2	
Beep Lat.:					
800 ms	452 (2.5%)	494 (1.7%)	448 (2.2%)	465 (2.8%)	465 (2.3%)
1200 ms	443 (1.7%)	476 (1.7%)	443 (2.1%)	439 (1.6%)	450 (1.8%)
1400 ms	440 (2.5%)	485 (0.7%)	435 (2.3%)	445 (1.7%)	452 (1.8%)
Mean Lat. (ms)	445 (6.6%)	485 (4.2%)	442 (6.6%)	449 (6.2%)	

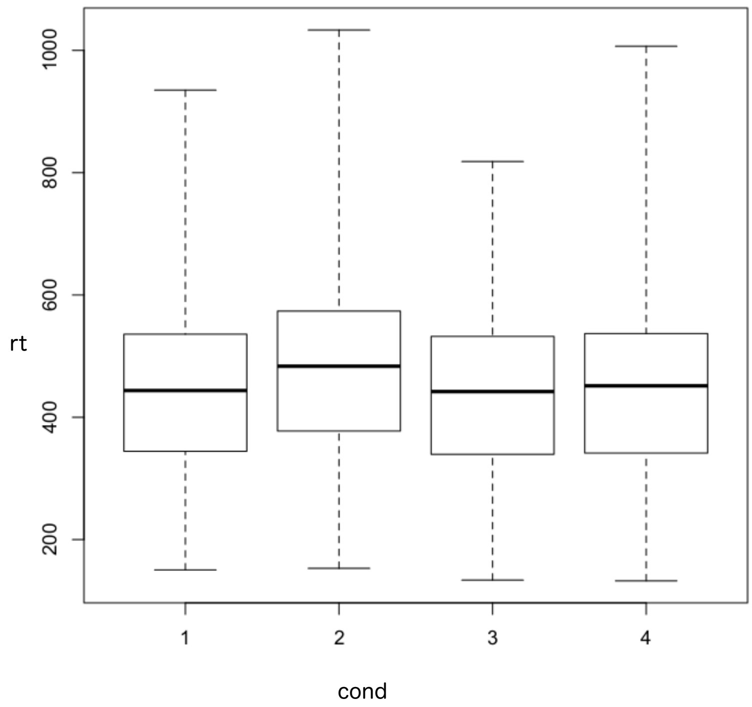


Figure 5.3: Mean Naming Latencies (RT) by Condition: Experiment 1

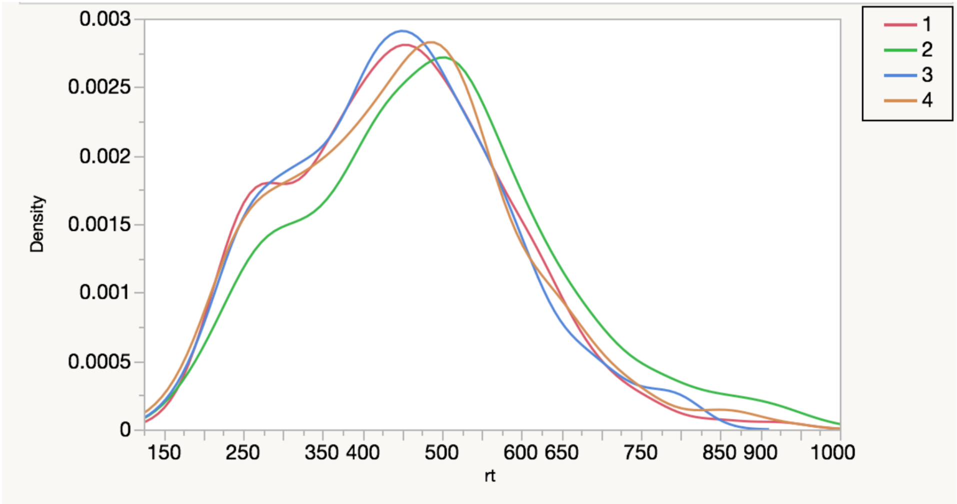


Figure 5.4: Densities of RT by Condition: Experiment 1

5.2.5.2 Errors

As part of this investigation, an error analysis was carried out: errors were categorised as "time out" (the subject said nothing), "voice key error", "disfluency" (e.g. stuttering), "stress error" (incorrect prosody), and "wrong item". The error rates were analysed proportionally using a generalised linear model (GLM) with a binomial distribution (link= logit). The analysis revealed no main effect of preparation time (χ^2= 6.41, p= 0.17) on error rate. While these speakers did not make frequent errors, disfluency errors were the most common, followed by wrong items and time outs (see Figure 5.9 below). Condition 2 contained the lowest number of errors (34 total), while the other conditions contained similarly-higher numbers. However, condition type did not have a significant main effect on error rates (χ^2= 2.02, p= 0.73). Nor was there was a main effect of block (χ^2= o.05, p=0.81).

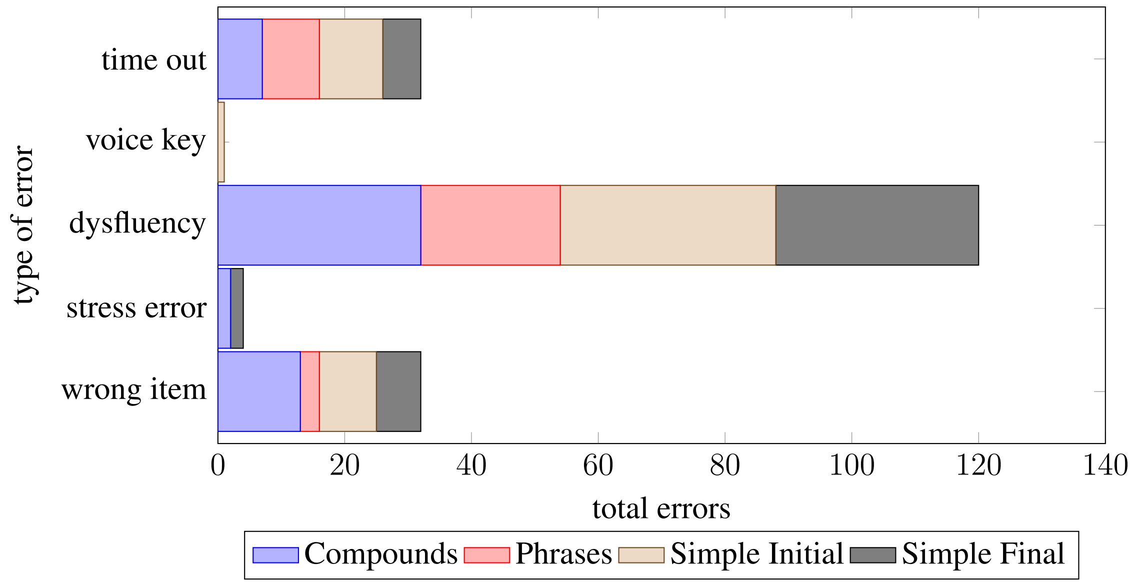


Figure 5.5: Total Errors: Experiment 1

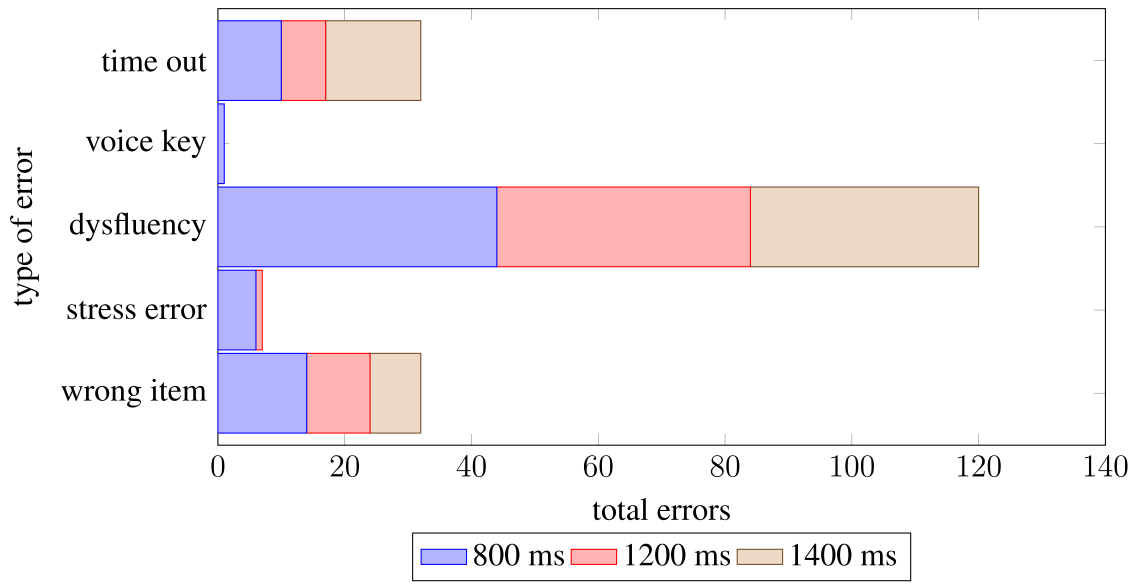


Figure 5.6: Error by Preparation Time: Experiment 1

5.2.5.3 Frequency Analyses and Word Length

For this analysis, the RT data were again submitted to a linear mixed models analysis, in which frequency, condition, and preparation time (beep) were all treated as the fixed effect factors. Once again, the maximally-appropriate structure that converged presented random intercepts and slopes for subject factors and a random intercept only for the by-items factor.[5] The models converged.

Neither total CELEX (Estimate= 0.08376, SE= 0.17255, t= 0.485) nor log (Estimate= 0.1264, SE= 0.2150, t= 0.588) frequency showed an effect in an interaction with condition. Neither simple CELEX frequency (Estimate= -0.2205, SE= 0.3713, t= -0.594) nor simple log frequency (Estimate= -11.217, SE= 8.344, t= -1.344) had an effect on the naming of compounds. Nor did the left morpheme frequency (Estimate= -5.161, SE= 6.971, t= -0.740), or the right (Estimate= 4.381, SE= 3.956, t= 1.107). Because only 5 different adjectives were used in the adjective-noun target words, the model was rank-deficient for left morpheme frequency effect. The frequency of the second morpheme of the adjective-noun phrase did not exhibit any effect on naming latencies (Estimate= -1.394, SE = 4.043, t= -0.345). Lastly, an analysis of the effect of word length (number of letters) resulted in no significant values for any of the lengths (all ts < 1.6).

5.2.6 Discussion

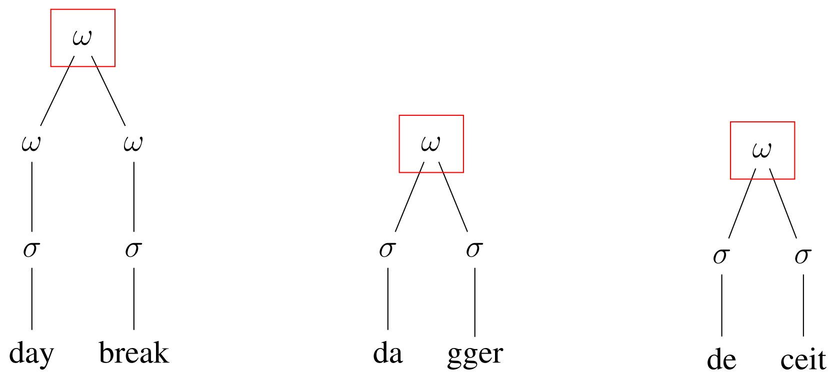
The results from Experiment 1 indicate that, when speakers had time to plan their utterances, responses containing compounds showed no difference in planning time from disyllabic, morphosyntactically-simple words. The naming latencies for adjective-noun phrases, conversely, were significantly longer than all other conditions. A cross-analysis of the naming latencies of compounds with both types of simple disyllabic words did not result in any significant difference (all ts < 2) , suggesting that speakers planned all three conditions similarly. Reaction times for the shortest preparation time (800 ms) were

[5] Formula: rt $\sim$ cond* freq + beep + (1 + cond | sub) + (1 | item).

generally slowest, but there was no effect of preparation time on latencies in a pairwise comparison. The size of the effect of condition type in Experiment 1 is similar to (if not greater than) that found in other experiments using this delayed naming paradigm (e.g. Wheeldon & Lahiri, 2002): the difference between compound and phrasal naming latencies in Experiment 1 was an average of 40 ms.

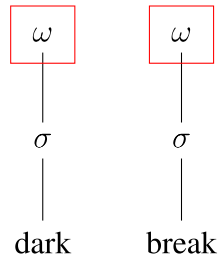
This is also consistent with our theories regarding the prosodic structure of compound words: the similarity in naming latencies of compound words to the morphologically-simple words suggests that speakers are planning these items as single prosodic structures, as in example 5.3:

(5.3) Units predicting naming latencies in compounds and simple words



Although the compound target items contained the same number of syllables and lexical words as the phrases, they elicited significantly different naming latencies. This fact leads us to venture that speakers are treating phrases as two discrete phonological units as in example 5.4:

(5.4) Units predicting naming latencies in adjective-noun phrases



Returning to the predictions made by the production models, we look at the processing of compounds in WEAVER++:

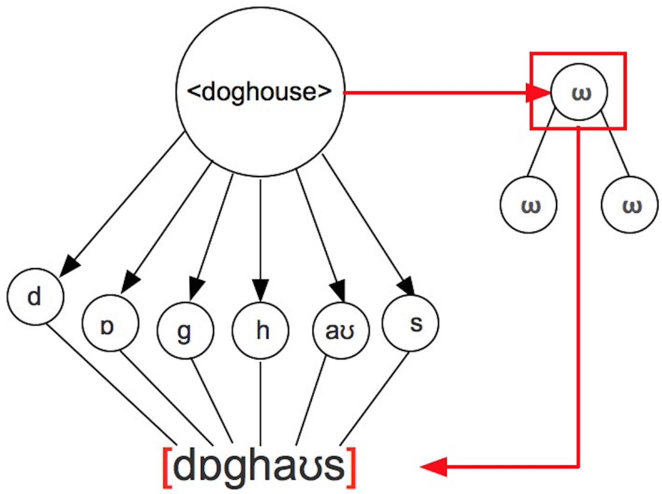


Figure 5.7: Encoding of "doghouse" in WEAVER++

In WEAVER++, Levelt et al. (1999) assume that lexicalisation of morphologically-complex words involves different routes: that is, newly-formed compounds are generated by different routes than frequently-used compounds. Therefore, "lexicalised" compound words such as *blackboard* contain a single lemma node, but multiple form nodes for each morpheme in the compound. Following this, the lexical representation (<doghouse>) enters the phonological encoding stage by accessing the word's default prosodic frame: for the compound words, this is a recursive Phonological Word that consists of two levels of structure: two Phonological Words within a larger Phonological Word constituent. At the same time, the Formulator accesses the word's segments. These are then placed within the prosodic frame(s) built not on the internal compound structure, but the external, recursive Phonological Word structure.

Errors have been used as evidence for the segmental organisation of sounds in speech production since the inception of modelling: many errors are seen as evidence of incremental encoding, of words and utterances, from left to right. In WEAVER++, phonological errors (that is, errors not related to the conceptual stage of production) were treated as a result of mapping processes, or "indexing failures of the device that maps syllabified phonological representations onto a mental syllabary" (Roelofs, 1997: 270). Errors in Experiment 1 were generally low, making up just 5.9% of the data. Compounds and initially-stressed words generated the most errors, as did the shortest preparation time (800 ms). Here we see that the majority of errors are of a phonological type: the most common errors made in Experiment 1 were disfluencies (e.g. repetitions, self-corrections, and noticeable pauses). Repetitions and self-corrections can be seen as failures that occur when the mapping process in WEAVER++ fails, or if multiple segment and syllable nodes share the same features (Roelofs, 1997: 271).

Finally, a word on frequency analyses. In this experiment, we found no significant effects of simple compound frequencies on reaction times; moreover, neither the first nor the second morpheme affected speed of response individually. Additionally, there was no evidence that the frequency of the adjective or noun in the phrasal conditions affected naming latency. Crucially for this experiment, there was no evidence of interaction of frequency and condition on reaction times: so while certain compounds elicited faster reaction times than others, the differences in the mean latencies across all conditions were not affected by word frequency. This is encouraging, as it suggests that we were able to control for word frequency in this task.

Our first test of the theories set out by the recent production models and prosodic framework has proven fruitful. We have elicited supportive evidence for the shape of the phonological encoding unit in English; our compounds, like those in Wheeldon & Lahiri (2002) generated similar naming latencies to morphosyntactically-simple words. However, we also know from Wheeldon & Lahiri (1997) that task conditions can affect naming latencies: their online task generated very different results than the delayed tasks. This leads us to consider what will happen when planning time is taken away: will compounds still be treated as single prosodic units, or will the removal of preparation time cause them to be planned differently?

5.3 Experiment 2: Online Production of Compounds

This experiment presents the stimuli of Experiment 1 in online task conditions. Online tasks involve the concurrent recall of verbal material without any opportunity to rehearse utterances. They are associated with an increased mental load, and consequently, an increase in errors. The most common type of online task is the picture-word interference paradigm, where a speaker is shown a picture and is asked to name it as quickly as possible, while at the same time hearing (or seeing) another (related or unrelated) word. In this experiment, speakers will be asked to name the target stimuli as quickly as possible, removing any time to plan ahead.

5.3.1 Hypotheses

Based on evidence we have from experimental online tasks, as well as the information we have gained from Experiment 1, we predict the following:

1. The process that prepares prosodic frames for utterances will be affected by loss of planning time, and this will be evident from the naming latency data.

This hypothesis is based on the claim that the unit of planning is flexible (Schriefers & Teruel, 1999; Smith & Wheeldon, 1999; Ferreira & Swets, 2002). Results from interference tasks containing phonologically-related distractors have suggested that the scope of the phonological encoding process is smaller than that of the grammatical encoding process (Schriefers, 1992, 1993; Meyer, 1996; Miozzo & Caramazza, 1999). The amount of planning depends on a number of variables (such as utterance complexity, working memory load, and time pressure), but allows for the possibility that speakers may be able to encode more than a single Phonological Word at the same time. However, as we saw in Chapter 2, there is uncertainty about *how* much smaller this scope actually is. Some theories have argued that only a small amount of advance planning occurs prior to articulation (MacKay, 1987; Jordan, 1990), while more recent theories hold that speakers are able to buffer a larger amount prior to articulation (Wheeldon & Lahiri, 1997; Levelt et al., 1999; Lahiri & Plank, 2010). Meyer (1996) found that naming latencies were only longer when a distractor word was phonologically-related to the first noun in the utterance; not the second. Schriefers & Teruel (1999) found that naming latencies were shorter *only* when a distractor word was related to the first word of an adjective-noun phrase in a picture-word interference task. More recently, Lange & Laganaro (2014) ran an online picture-word interference production experiment using French phrases, including adjective-noun phrases (e.g. *grand chat*). They found that only the first element of the phrase was primed by a phonologically-related distractor; no priming effect was found when the second element in the phrase was primed.

Planning scope can be connected to working (short-term) memory (Ellis, 1980; Levelt, 1989; Bock, 1996; Martin & Freedman, 2001; Jones, Macken, & Nicholls, 2004; Jacquemot & Scott, 2006; Page et al., 2007; Wagner et al., 2010). The underlying representations generated at each stage in the production process are assumed to be maintained by working memory, with the output from each process being stored in limited-capacity buffers at each level (Martin & Freedman, 2001). During phonological encoding, the unfolding phonological segments are compiled and stored in a buffer that contains the phonological output forms (see Figure 5.8 below).

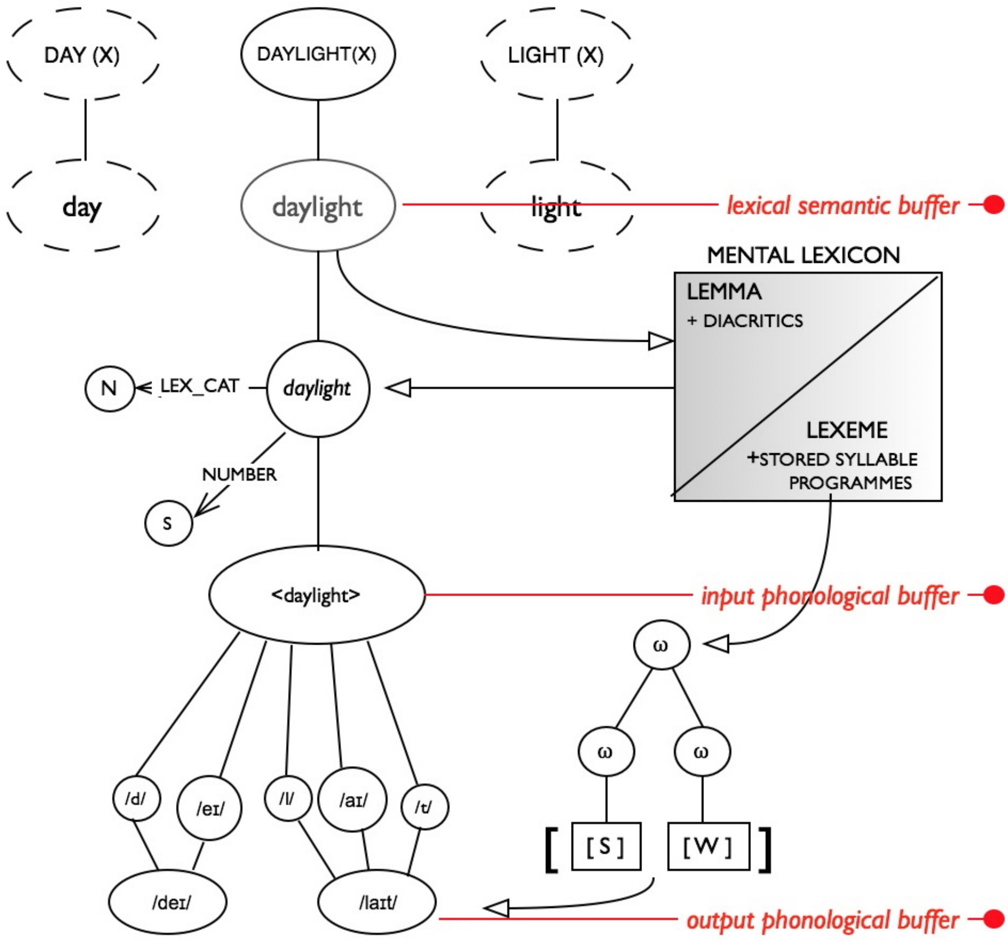
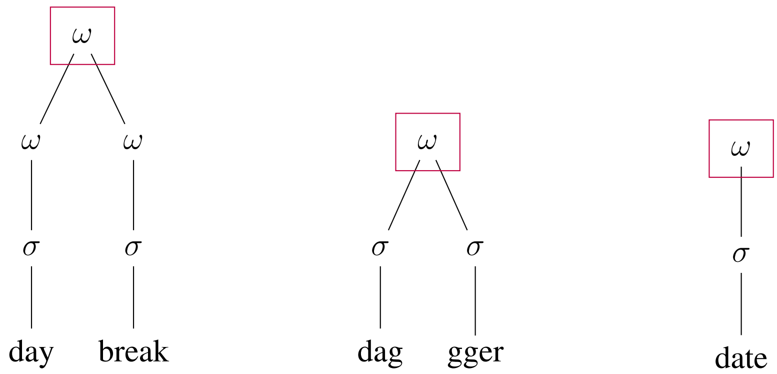


Figure 5.8: Working Memory Buffers in Language Production

When planning time is removed, the amount of information the Formulator is able to encode is limited. WEAVER++ assumes an incremental encoding process, in which structures are prepared from left to right. Following this, we hypothesise that the removal of planning time will reduce the number of prosodic frames the Formulator is able to prepare; however, crucially, these frames will still be built according to phonological, not lexical words.

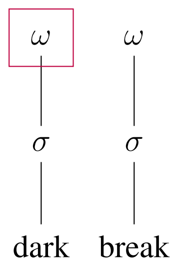
2. Compounds will retain their recursive phonological structure, and elicit naming latencies similar to those of disyllabic morphosyntactically-simple words.

Our next hypothesis addresses the structure of compound words. If prosodic frames are still built according to Phonological Words, then it follows that compounds will retain their recursive phonological structure and be treated once again as single prosodic units. Therefore, compounds should elicit significantly longer naming latencies, because the PWd unit is more complex than the initial constituent of the phrases.



Conversely, phrases will elicit shorter naming latencies than compounds and disyllabic words, because they are built from two discrete Phonological Words:

(5.5) Predicted planning units for adjective-noun phrases



If compounds retain their prosodic shape even when planning time is removed, this will bolster our claims regarding the shape of the planning unit at this stage.

5.3.2 Stimulus Selection

The aim of Experiment 2 was to test whether removing the preparation time of Experiment 1 would affect post-lexical encoding of compound words. Because evidence suggests that only the first unit of the target will determine the production latency, we had to ensure that the control conditions contained corresponding structures. Consequently, a new condition was added to the target wordlist: monosyllabic simple words replaced the final-stressed disyllabic words of Experiment 1. The rest of the target items remained the same. 15 new filler words were added.

Table 5.7: Stimuli: Experiment 2

Condition 1: Compounds	**Condition 2: Phrases**	**Condition 3: Disyllabic**	**Condition 4: Monosyllabic**
daybreak	dark break	denim	dame
daytime	dark time	debit	date
daylight	dark light	dagger	daze
groundhog	green hog	gravel	grease
groundwork	green work	griddle	greed
groundnut	green nut	gravy	greet
nightcap	nice cap	nibble	nine
nightgown	nice gown	nickel	Nile
nightmare	nice mare	nitrate	ninth
lockjaw	long jaw	locker	lob
locksmith	long smith	locust	log
locknut	long nut	lodger	loft
bathrobe	bad robe	banter	bag
bathroom	bad room	ballad	band
bathtub	bad tub	basil	bat

Again, word familiarity was confirmed by rating data from 35 participants using a native speaker judgement task. These participants were the same as in Experiment 1. Mean word familiarity for the new condition was high (4.942). The mean judgement ratings for each condition did not significantly differ from each other ($p=0.326$). The new rating data is included alongside the data from Experiment 1 in Table 5.8.

Table 5.8: Weighted Averages of Word Familiarity: Experiment 2

Cond 1:	Score	**Cond 2:**	Score	**Cond 3:**	Score	**Cond 4:**	Score
daybreak	5	dark break	5	denim	4.97	dame	4.81
daytime	5	dark time	5	debit	4.84	date	5
daylight	5	dark light	5	dagger	5	daze	5
groundhog	4.84	green hog	5	gravel	5	grease	5
groundwork	4.88	green work	5	griddle	4.91	greed	4.97
groundnut	2.88	green nut	5	gravy	5	greet	5
nightcap	4.91	nice cap	5	nibble	4.84	nine	5
nightgown	5	nice gown	5	nickel	4.84	Nile	4.88
nightmare	5	nice mare	4.91	nitrate	4.31	ninth	5
lockjaw	4.81	long jaw	5	locker	5	lob	4.66
locksmith	4.97	long smith	4.78	locust	4.91	log	5
locknut	2.84	long nut	5	lodger	4.44	loft	4.81
bathrobe	5	bad robe	5	banter	4.94	bag	5
bathroom	5	bad room	5	ballad	4.97	band	5
bathtub	5	bad tub	5	basil	5	bat	5
Mean Score:	4.675		4.979		4.866		4.942

Table 5.9: Mean Variables: Experiment 2

	Comp 1	**Comp 2**	**Comp**	**Adj**	**N**	**Disyll**	**Mono**
CELEX Freq	291.5	258.0	5.3	298.7	259.0	2.9	19.6
Log Freq	2.1	1.6	0.5	2.4	1.6	0.5	1.0
No of Letters	4.4	3.8	8.2	4	3.8	5.9	4.1

5.3.3 Design

The experiment was designed using E-Prime 2.0 software. There were 5 blocks of 15 trials (75 items total including filler words) with optional breaks between the blocks. Items were distributed pseudo-randomly within the blocks, and three different versions of the experiment were presented. Experimental stimuli were only presented once in this task- this was done to avoid word repetition effects (cf. Humphreys et al. (1990)).

5.3.4 Participants

Twenty participants between the ages of 18 and 26 took part in Experiment 2. Participants were monolingual speakers of Southern British English. They were recruited through mailing lists at the University of Oxford and were compensated for their participation. They were tested individually in a soundproofed booth.

5.3.5 Procedure

Because the task removed planning time, some alterations had to be made to both the stimuli and intended utterance. If the naming latencies in this task reflect the first available Phonological Word, then the intended utterance used in Experiment 1 ("It was...") would be expected to generate flat results. Therefore, the intended response was simplified: speakers were told to produce only the target word. There were no preparation beeps, nor was there a deadline beep: speakers were asked to answer immediately after hearing the auditory cue: *"What was it?"* In order to avoid cue-jumping, the participants were asked to answer only **after** the question was asked. Latencies were measured from the end of the final [t] plosive of the auditory prompt to the onset of the target's utterance.

5.3.6 Results

5.3.6.1 Latencies Analysis

Measurements were undertaken in the same way as in Experiment 1. Errors resulted in a loss of 9.5% of data. Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error. All data points beyond two standard deviations from the mean were counted as outliers and removed. This, combined with the errors, resulted in a total loss of 12.4% of data. Again, the target *nightmare* was removed from analysis.

The RT data were submitted to a linear mixed models analysis, in which naming latency (RT) was modelled as a function of the fixed effect factor, condition type (cond). Subjects and items were treated as random factors. The maximal model had condition as the fixed effect, random slopes for condition in the by-subject analysis, and random slopes for condition in a by-item analysis. This model[6] failed to converge. The next maximal model to converge contained random slopes and intercepts for the condition in the by-subject analysis, and random intercepts in the by-item analysis.[7] This model converged and its output is shown in Table 5.10. The model was comfortably homoscedastic, as evidenced by the diagnostic plots in Figure 5.9. Mean reaction times and percentage error rates for Experiment 2 are shown in Table 5.11. The mean latencies for compounds and disyllabic words were statistically similar to one another, as were the latencies for adjective-noun phrases and monosyllabic words. Furthermore, compounds and disyllabic words differed significantly from phrases and monosyllabic words. Notably, there was an effect of block on RT: naming latencies increased significantly in the fifth (Estimate= 41.087, SE= 12.548, t= 3.274*) and final experimental block (Figure 5.12).

[6] mla.maxstar <- lmer(rt ~ cond + (1+cond|sub)+ (cond|item)).
[7] mla.maxplus3 <- lmer(rt ~ cond + (1+cond|sub)+ (1|item))

Table 5.10: Linear Mixed-Effects Analyses of RT Data in Experiment 2

	Estimate	**SE**	**t Score**
All Conditions:			
(Intercept)	263.603	15.653	16.840
Adj-Noun Phrases	-35.053	9.542	-3.674*
Disyllabic	-2.127	8.066	-0.264
Monosyllabic	-34.452	9.691	-3.555*
Pairwise Comparisons:			
Adj-N Phrases as Intercept:			
(Intercept)	228.5501	17.3755	13.154
Compounds	35.0531	9.5415	3.674*
Disyllabic	32.9260	8.8631	3.715*
Monosyllabic	0.6012	8.8258	0.068
Disyllabic as Intercept:			
(Intercept)	261.476	16.681	15.675
Compounds	2.127	8.066	0.264
Adj-Noun Phrases	-32.926	8.863	-3.715*
Monosyllabic	-32.325	7.187	-4.498*
Monosyllabic as Intercept:			
(Intercept)	229.1512	14.7936	15.490
Compounds	34.4519	9.6915	3.555*
Adj-Noun Phrases	-0.6012	8.8258	-0.068
Disyllabic	32.3248	7.1869	4.498*

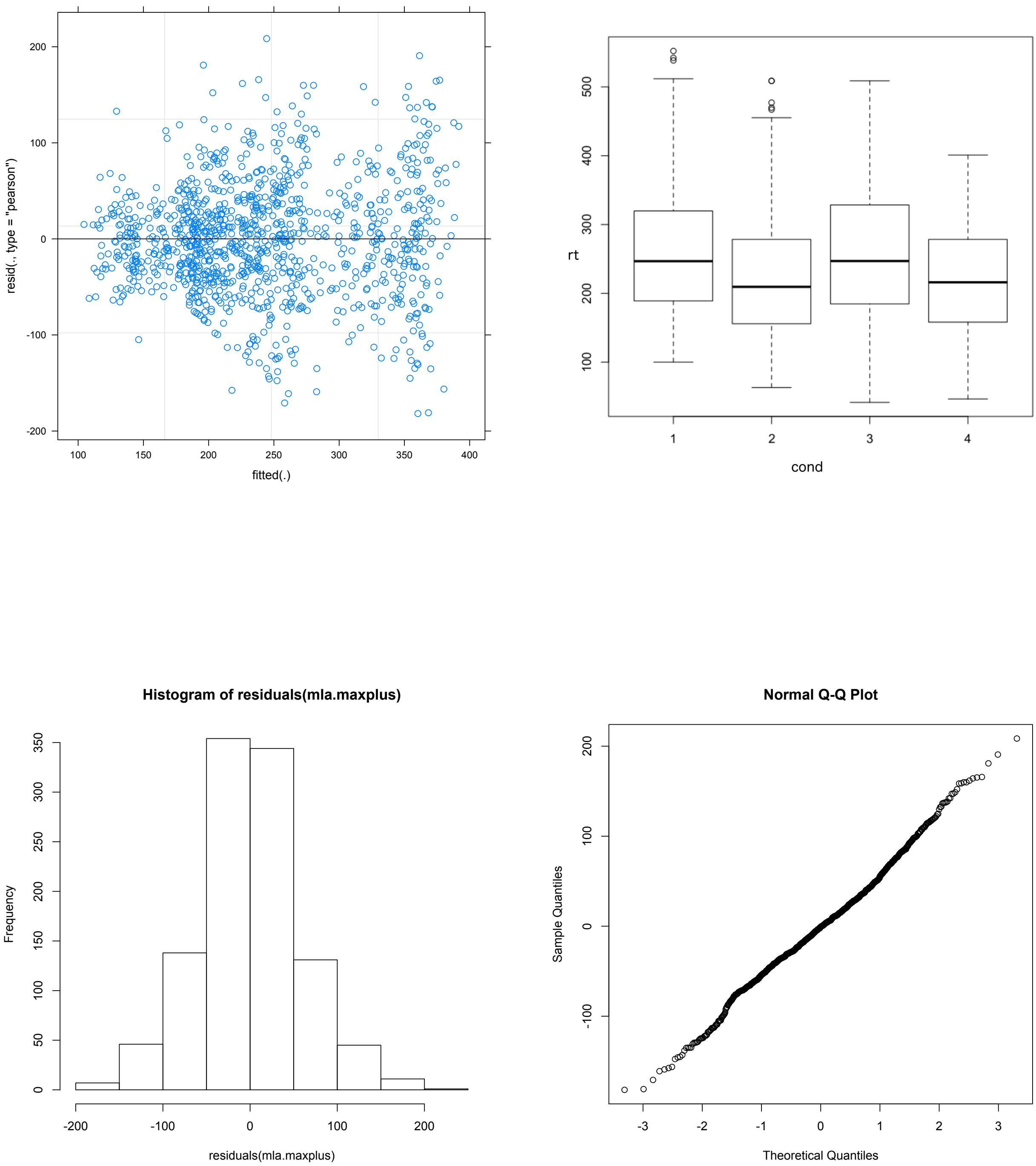


Figure 5.9: Residuals Analyses: Experiment 2

Table 5.11: Mean Naming Latencies (in ms): Experiment 2

	(1) Compounds	(2) Phrases	(3) Disyll	(4) Mono
PWds	1	2	1	1
LexWds	2	2	1	1
Syllables	2	1	2	1
rt (ms)	262 (8.3%)	223 (8.6%)	259 (7.0%)	221 (14.3%)

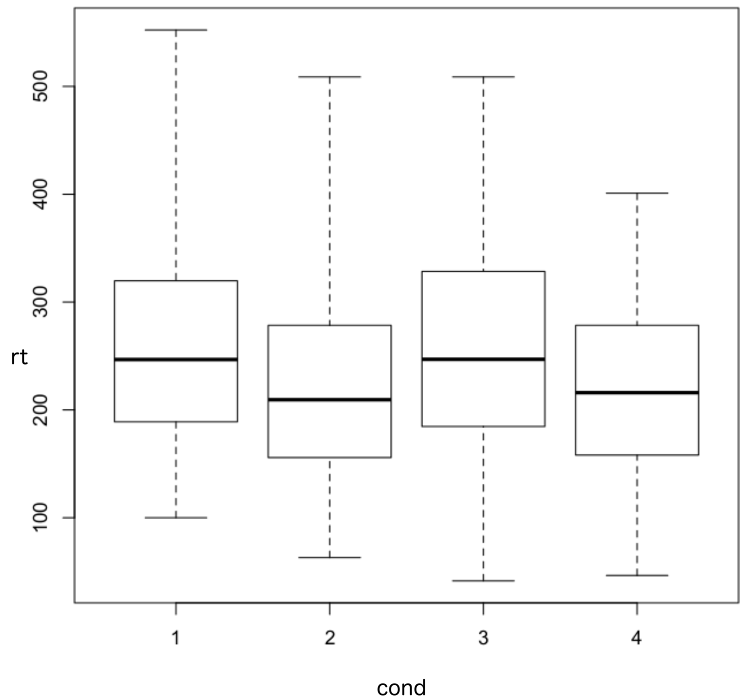


Figure 5.10: Mean RT by Condition: Experiment 2

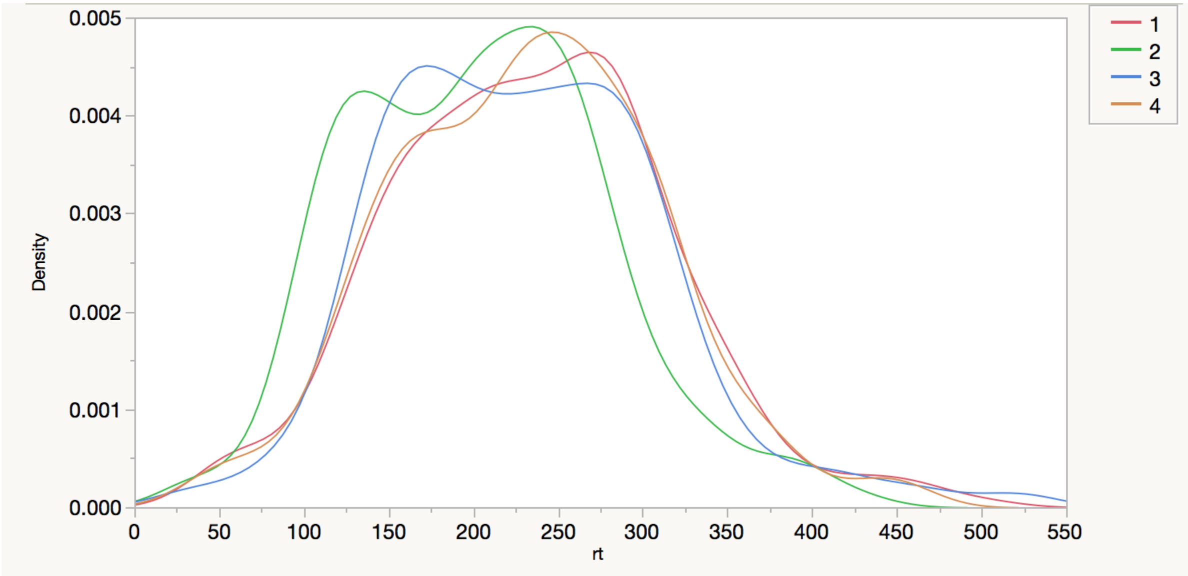


Figure 5.11: Densities of RT by Condition: Experiment 2

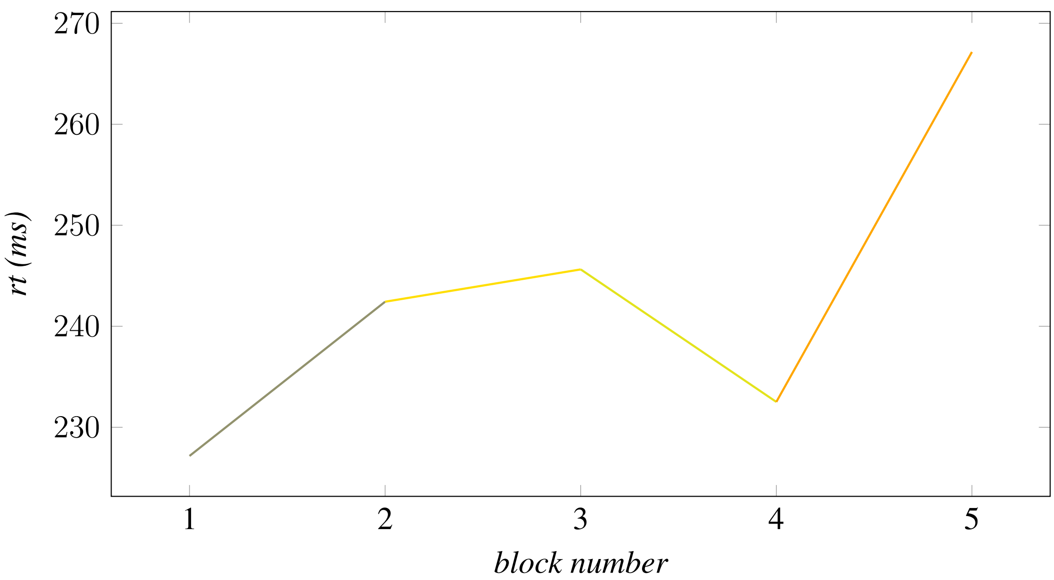


Figure 5.12: Mean RT by Block: Experiment 2

5.3.6.2 Errors

Disfluencies were the most common error for speakers to make in this task, followed by null responses. The ratio of errors to trials was higher in this task than in Experiment 1. Errors of incorrect item and incorrect prosody were fairly infrequent. The error rates were analysed proportionally using a generalised linear model (GLM) with a binomial distribution (link= logit). The analysis revealed no main effect of condition (χ^2= 0.50, p= <. 97) on error rate. While these speakers did not make frequent errors, disfluency errors were the most common, followed by wrong items and time outs (see Figure 5.13 below). We found that the number of errors (Figure 5.14) decreased in blocks 2 and 3, and rose in blocks 4 and 5; however no main effect of block was found in the analysis (χ^2= 2.83, p= 0.09).

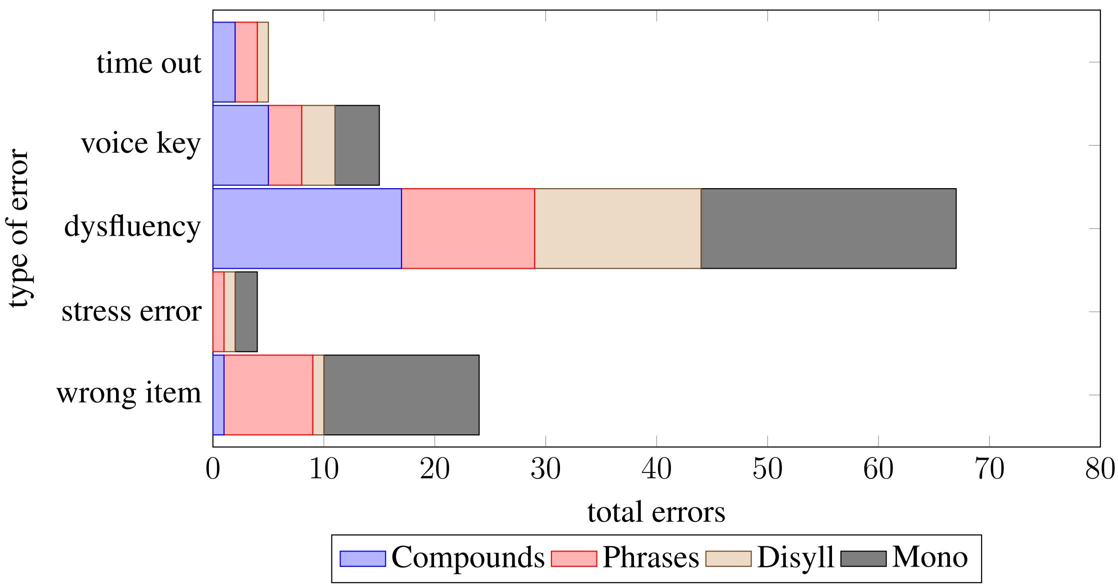


Figure 5.13: Total Errors: Experiment 2

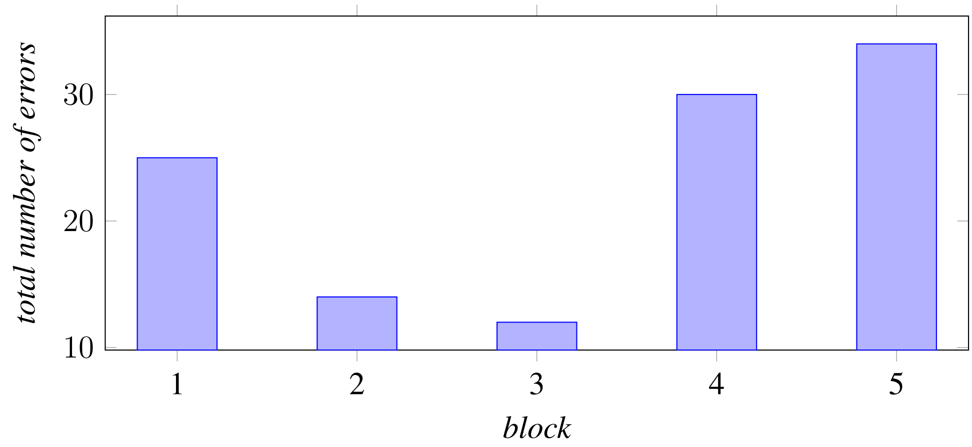


Figure 5.14: Errors by Block: Experiment 2

5.3.6.3 Frequency Analyses

The RT data were again submitted to a linear mixed models analysis, in which frequency and condition were treated as the fixed effect factors. Once again, the maximally-appropriate structure that converged presented random intercepts and slopes for condition in the by-subject factors and a random intercept only for the by-items factor.[8] The model converged. Neither log frequency (Estimate= -0.6829, SE= 1.8007, t= -0.379) nor CELEX frequency (Estimate= -0.06088, SE= 0.05448, t= -1.117) showed an effect for an interaction with condition. Nor was there was any overall effect of frequency on naming latencies: log (Estimate= 6.099, SE= 6.415, t= 0.951) or CELEX (Estimate= 0.01699, SE= 0.17199, t= 0.099). In the compound frequency analysis, neither the first (Estimate= 6.295, SE= 11.801, t= 0.533) nor second (Estimate=10.557, SE= 5.552, t= 1.902) morpheme frequency reached significance for an interaction with condition type on reaction times (however, the second morpheme frequency came close). The first morpheme of the adjective-noun phrase had no effect (Estimate= 2.636, SE= 15.469, t= 0.170), while the noun was similar to the second compound constituent (Estimate= 8.245, SE= 4.319, t= 1.909). Lastly, word length did not appear to have any effect (all ts < 1.6) on naming latencies; this result will be discussed below.

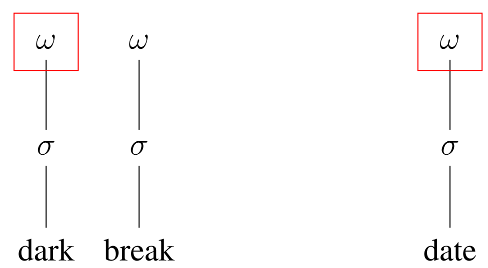
[8]Formula: rt ∼ cond + freq + (1 + cond | sub) + (1 | item).

5.3.7 Discussion

As in Experiment 1, the results from Experiment 2 showed a significant main effect of Condition on reaction times: responses containing compounds also showed no difference in naming latencies from disyllabic, morphosyntactically-simple words. In contrast to Experiment 1, however, speakers in this task consistently took **longer** to produce utterances containing compounds (262 ms) than they took to produce the adjective-noun phrases (223 ms). Furthermore, while a cross-analysis of the naming latencies of compounds and disyllabic words did not result in any significant difference (t= -0.264), it did show a significant difference when comparing reaction times for compounds to both the adjective-noun phrase (t= -3.674*) and monosyllabic simple word conditions (t= -3.555*). Adjective-noun phrases, conversely, had statistically-similar naming latencies to the monosyllabic words (t= 0.068).

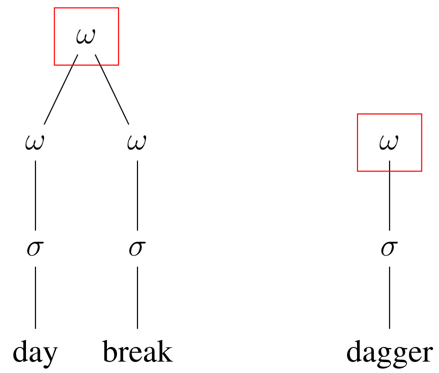
These results are consistent with those found in Meyer & Schriefers (1991), Meyer (1996), Wheeldon & Lahiri (1997), Schriefers & Teruel (1999), and Lange & Laganaro (2014), where only the first element of adjective-noun phrases received priming. The mean naming latency (223 ms) for adjective-noun phrases reflects encoding of the first phonological unit, resulting in similar naming latencies to the monosyllabic word condition (221 ms): this suggests that, when planning time was removed, speakers were only able to encode the first unit of the adjective-noun phrase condition.

(5.6) Restricted planning scope in Experiment 2: Phrasal and monosyllabic word conditions



More revealing, however, were the reaction times for compound words: compounds took significantly longer to produce than phrases (approximately 40 ms). As in Experiment 1, this indicated that phrases and compounds were being planned differently. The RT data for compounds suggests that speakers waited until the entire phonological unit was ready before beginning articulation: for phrases, this was simply the first unit (the adjective). For compounds, however, this was the entire compound structure.

(5.7) Planning of compounds and disyllabic words in online task conditions



These findings are in line with the theory that reaction times for these experiments are based on prosodic, not lexical words. Importantly, speakers do not plan the parts of the compounds as separate lexical words for the purpose of phonological encoding: they wait until the entire prosodic unit is ready, then initiate speech.

We can relate this to what happens in the phonological encoding process. In Experiment 1, the entire phrase received activation when speakers had time to plan ahead. Here, however, speakers are only able to access the first default prosodic structure of the phrase:

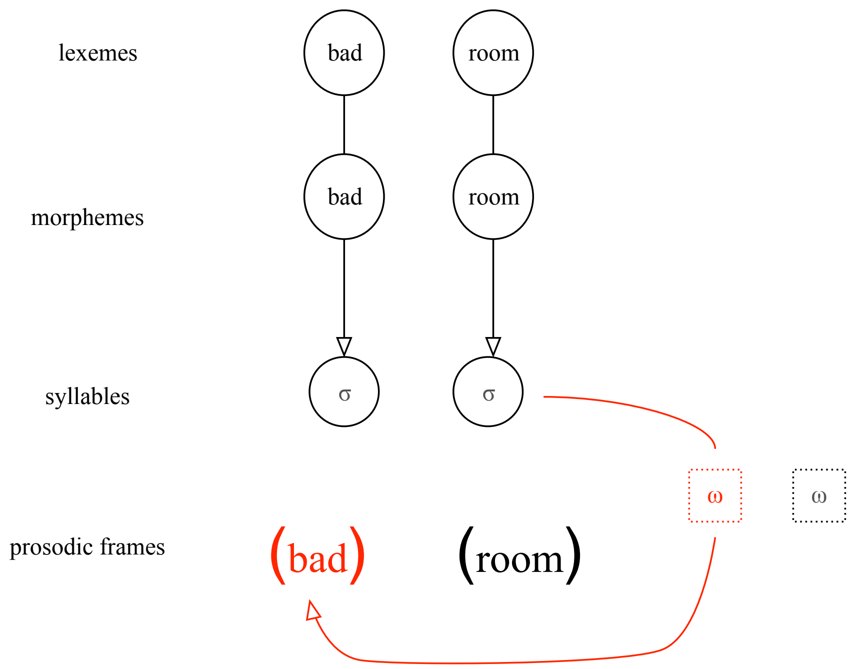


Figure 5.15: Encoding "bad room" in Online Task Conditions

The difficulties associated with online planning tasks extend beyond limited planning scope. Although naming latencies in this task were (understandably) shorter than Experiment 1, it also resulted in a higher number of errors. This effect is regularly seen with online tasks (Baddeley, 1966; Luck & Vogel, 1997; Oomen & Postma, 2001; Unsworth & Engle, 2007; Rouder et al., 2011, Cholin et al., 2011).[9] Errors rose significantly in the final two blocks, although the experiment only lasted about 20 minutes. Disfluencies remained the most common error, followed by wrong item, and interestingly, the monosyllabic word condition resulted in the highest number of errors overall (14.3%). Despite this, errors did not have a significant effect on the results. In

[9]Further evidence can also be found in Event-Related Potential (ERP) studies focusing on the Error-Related Negativity (ERN) component. This component is regularly associated with the error-monitoring system, showing sensitivity about 80 ms after incorrect responses (Bernstein et al., 1995; Holroyd & Yeung, 2003). In Ganushchak & Schiller (2008), the researchers found that the amplitude of the ERN decreased under time pressure, suggesting that this component was less able to monitor for errors when limited. Additionally, they found an increase in errors under time pressure.

addition to containing more errors, the final experimental block was also associated with significantly-slower naming latencies (t=3.274*). These findings are all in line with the mental load associated with online task conditions.

In the discussion for Experiment 1, we talked about how errors are often used as evidence for the incremental planning of speech. WEAVER++ assumes an encoding process that works from left to right, "like in weaving a fabric" (Roelofs, 2008). Evidence from implicit priming tasks (such as those used to test priming of only one unit of adjective-noun phrases) also lend support towards this process: priming of the first syllable speeds up production of a disyllabic word, but not priming of a second. When planning time is restricted as it was in Experiment 2, the encoder, moving from left to right, activates the first phonological frame: in phrases (cf. example 5.6), this is the adjective *dark*. In compounds, however, the first phonological frame is not the first lexical word of the compound; it is the recursive Phonological Word that makes up the entire compound (cf. example 5.7).

The incremental, left-to-right generation of speech has implications for the prosodic unit in this task: Wheeldon and Lahiri (1997) ascribe the difference in naming latencies in their online task to the "complexity" of the first prosodic unit. They refer to the difficulty associated with generating surface syntactic structure and grammatical gender as indications of this complexity, so the phonological form of *ik zoek het* required more work than *ik zoek*). This has some interesting implications for the conditions in this task. Compound words are single morphosyntactic units ($[\text{doghouse}]_{\text{M}}$), contain two Phonological Words ($(\text{dog})_{\omega}$ $(\text{house})_{\omega}$), but also act as a single Phonological Word as evidenced by the results above ($(\text{doghouse})_{\omega}$). This lends evidence to the idea that lexicalised compounds are entering the phonological encoding process already specified for shape; at no point in post-lexical process are they treated as individual Phonological Words.

Lastly, a word on frequency and length results. While there was no sign of an interaction between frequency and condition on reaction times, there also seemed to be no effect whatsoever of frequency on naming latency. Word length also appeared to have no effect on reaction times in this task, something that is contradictory to the findings of many studies on word length and naming (cf. Klapp and Erwin, 1976). This is also important when we refer to the incremental encoding models such as WEAVER++. Meyer et al. (2003) have a possible explanation for this: "the length effect disappears if responses are initiated as soon as one syllable program has been retrieved even if the complete phonological representation is generated before speech output" (145). That is, when the deadline to initiate speech was such that articulatory planning for longer words (e.g. compounds and disyllabic words) could not be completed, word lengths effect would be nullified.

5.4 General Discussion

In the series of experiments reported above, we found that English native speakers treat two-word compounds as single units, with RT data statistically different from that of adjective-noun phrases. In this first set of experiments (Experiments 1 and 2), we compared production latencies of utterances containing compounds to those of utterances containing syntactic phrases (adjective-noun phrases) or morphologically simple words. In the delayed task, naming latencies reflected the total number of Phonological Words in the utterance. Speakers took an average of 40 ms longer to produce utterances containing syntactic phrases than noun-noun compounds. Utterances containing compounds generated statistically similar production latencies to morphologically-simple words. Conversely, when planning time was taken away in Experiment 2, latencies were only sensitive to the first phonological unit of the utterance. Both of these findings are in line with those in Wheeldon & Lahiri (1997, 2002).

In both Experiments 1 and 2, we saw a significant effect of condition type on production latencies. In the delayed task (Experiment 1), compound words elicited shorter naming latencies than adjective-noun phrases, similar to those of disyllabic morphosyntactically-simple words. In this experiment, the number of syllables was held constant across all conditions, while the number of Phonological Words and lexical words, and the placement of stress all varied. Despite these variations, the only thing that reliably predicted naming latencies was the total number of prosodic units in the utterance. Although compounds contained two morphosyntactic words (and therefore two Phonological Words), they were treated as single prosodic units in this task, generating nearly identical naming latencies to morphosyntactically-simple words. In the online task (Experiment 2), we saw a different pattern of effects: compound words elicited longer naming latencies than phrases, similar only to those found for disyllabic morphosyntactically-simple words. Adjective-noun phrases were associated with shorter latencies in this task, similar only to those found for the monosyllabic simple words.

Experiment 2 made clear the importance of testing the stimuli in both task types: when speakers had enough time to plan, the total number of prosodic units dictated the naming latencies. When planning time was restricted, however, the phonological encoding process was only able to prepare the first prosodic unit before articulation: this was evident in the naming latencies in Experiment 2. Crucially, speakers sacrificed utterance initiation time in order to access well-formed prosodic units: in the case of the compound words, they did not begin speaking until the entire compound structure was encoded. While the data in Experiment 1 serves to support existing evidence for the Phonological Word in phonological encoding, Experiment 2 offers an important additional contribution to this process: it revealed that speakers still preferred to plan compounds as a single prosodic structure. These findings support both the theories regarding the prosodic word structure of compounds and the shape of the planning unit in the word-form encoding process.

The secondary aim of these experiments was to establish a foundation for the rest of this study. These results indicate that we have a viable paradigm for obtaining information about both the planning unit and the planning scope of phonological encoding. Now that we have established the planning unit in English, and found evidence for the treatment of compounds in the phonological encoding process, we will move forward by increasing the complexity of the target utterance. If speakers are indeed planning English compounds as single prosodic units, then it follows that they should continue to act as single prosodic units when cliticised. The following chapter will test the production of compounds and clitics together in order to ensure that compounds are indeed being planned as single prosodic units under these conditions.

6

The Phonological Encoding of Clitics by Native English Speakers

6.1 Introduction

Following the results of Experiments 1 and 2, this chapter aims to further test the claims made regarding the structure of the planning unit and the process by which it is planned during phonological encoding. We do this by increasing the complexity of the utterances: evidence from theoretical frameworks and experimental studies (including those in Chapter 5) predicts that, if compounds are truly acting as single prosodic units for the purposes of phonological encoding, then clitics will attach leftwards to the structure as a whole in English. Therefore, the experiments in this chapter test the behaviour of clitics with our target conditions. One of the key aims of this analysis will be to look at how compounds and clitics interact: that is, to investigate whether clitics attach to compound words to form a single prosodic unit, or whether they only exhibit some other behaviour (such as attaching to only one of the compound morphemes or pro-cliticising to the following prosodic unit).

The tasks in this chapter follow the same experimental designs of Experiments 1 and 2, altering only the auditory prompt and the intended utterance. First, Experiment

3 will look at how compound words are phonologically-encoded when followed by a clitic in delayed task conditions. Experiment 4 investigates this same process in an online task, similar to Experiment 2. The results of this online task will be important for the question of direction of cliticisation, because naming latencies in this task only reflect the first prosodic unit of the utterance, and will consequently indicate whether the clitic has attached to this unit or not. This chapter seeks to provide further evidence for the prosodic shape of compounds in native English speakers by eliciting reaction-time measurements for compounds and clitics together. Predictions based on these findings will be discussed at the end of the chapter.

6.2 Experiment 3: Delayed Production of Clitics

The experiment reported in this section was designed to elicit information about how speakers plan cliticised compounds when they have time to prepare their utterances. This experiment consists of a delayed production task with a question-answer paradigm.

6.2.1 Hypotheses

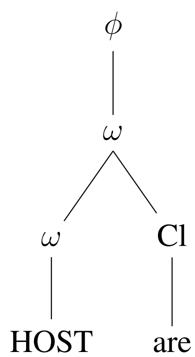
The primary aim of this experiment is to increase the complexity of the prosodic unit and elicit information on how it is planned during native English phonological encoding. Furthermore, we seek to gather data reflecting the direction of cliticisation in these utterances. In light of the results of Experiments 1 and 2, as well as the evidence given regarding the phonological behaviour of clitics, the main hypotheses tested in this study are as follows:

1. Despite an increase in complexity, the planning unit will remain the Phonological Word in this task.

This follows the hypotheses made in Experiment 1, namely that naming latencies will be predicted by the number of prosodic, not lexical, units in the utterance. This task will follow the same task conditions as Experiment 1, meaning that speakers will have time to plan their utterances. We saw in Chapter 2 that clitics often attach to neighbouring "host" words, forming new structures (e.g. *could have* > *could've*); we also saw that clitics display a wide range of behaviour during this attachment process (e.g. Zwicky's bound morphemes in §2.3.1).

The question we ask here is the following: for the purposes of the construction of prosodic frames, how will clitics in English behave? In this task, we invoke cliticisation by presenting experimental conditions in which an auxiliary verb "are" should cliticise leftwards to a target (host) word, as it would in regular (or rushed) speech. Based on the evidence from Experiment 1, as well as the results from Wheeldon and Lahiri (1997) (cf. §2.3.3.1), we predict that the prosodic structure of the target items will continue to dictate the size and shape of the planning unit (the Phonological Word). In their study, Wheeldon and Lahiri found strong evidence that clitics were being treated as weak elements by the Dutch speakers, reducing and attaching leftwards to Phonological Words. Following this, we predict that the results of this task will be similar: the auxiliary will reduce and cliticise leftwards to the prosodic unit. We predict that this unit will consist of the target word **plus** the clitic (e.g. *daggers* + *are*). We theorise that the target word will constitute its own Phonological Word, and the auxiliary *are* will become unstressed and attach to its host, forming a recursive Phonological Word:

(6.1) Cliticisation



This means that the prosodic frames being generated by the Formulator will only reflect the outer (recursive) phonological structure again. In this case, the frame will include the target word and the clitic.

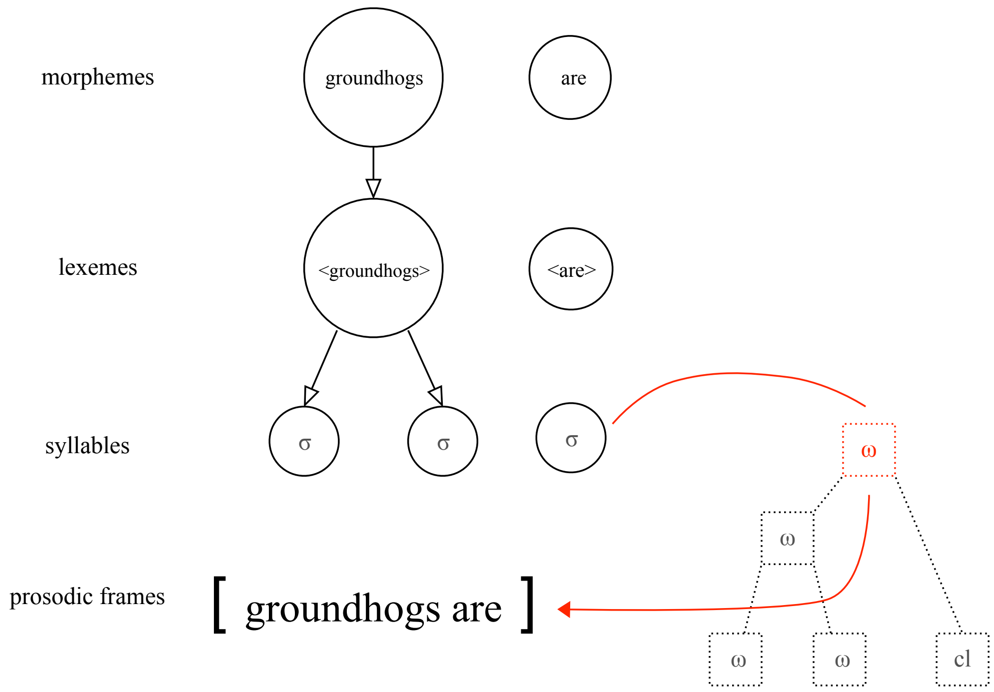


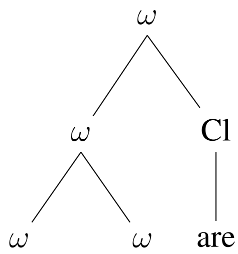
Figure 6.1: Encoding of "groundhogs are" in WEAVER++

Therefore, as in Wheeldon & Lahiri (1997), naming latencies in this task will only reflect the larger recursive Phonological Word structure. Once again this will be most visible in the compound word condition, where we maintain that compounds will continue to be planned as single prosodic units (Hypothesis 2).

2. **Naming latencies will indicate that compounds are still being planned as single prosodic units for the purposes of phonological encoding; furthermore, clitics will attach to the recursive structure of the compound, forming a larger recrusive prosodic constituent.**

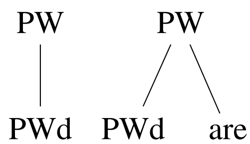
In Chapter 2, we theorised that clitics could attach to other Phonological Words, forming a larger recursive structure. In the case of the compound target words in our stimuli list, we predict that the auxiliary "are" will attach leftwards to the compound as a single prosodic unit. That is, the naming latencies of compounds will reflect the outmost Phonological Word in the structure, which will include the clitic (see Example 6.2):

(6.2) Compound+clitic structure



If the prosodic frames are being dictated by this recursive Phonological Word structure, then latencies for compounds will not differ from those of disyllabic morphosyntactically-simple word conditions. Phrases, on the other hand, will generate significantly different naming latencies from all other conditions in the task, as the total number of Phonological Words in the utterance will be greater than any other condition.

(6.3) Noun Phrase+clitic structure



In summary, we predict that in this delayed task, as in Experiment 1, it will be the total number of Phonological Words in the utterance that dictates the naming latencies.

6.2.2 Stimuli Selection

In order to ensure that speakers were generating well-formed sentences, we had to present target words that would easily cliticise with an auxiliary. Auxiliary reduction is widespread with English auxiliary verb forms such as *are, is, am*, particularly when following a plural subject (Labov, 1969). Therefore we produced a new target wordlist, as well as different auditory prompts (see Table 6.1). The experimental materials were constructed from 60 items, 15 per condition type. Four experimental conditions were distinguished: noun-noun compounds, adjective-noun phrases, initially-stressed disyllabic simple words, and monosyllabic simple words. All target items were presented in the plural, e.g. *dishcloths*, and the auditory prompts were designed to invoke the reduction and attachment of the auxiliary word *are* to these targets. The monosyllabic condition was chosen as a control for the adjective-noun phrase, because we hypothesise that the adjective will be treated as a separate prosodic unit for the purposes of planning.

As in Experiments 1 and 2, targets were arranged into five sets beginning with one of five phonemes: /d/, /g/, /n/, /l/, and /b/. These phonemes were chosen in order to more easily identify boundaries during analysis (cf. Rastle & Davis, 2002). 8 disyllabic and 8 monosyllabic filler words were included as well.

Table 6.1: Stimuli: Experiment 3

Condition 1: Compounds	**Condition 2: Phrases**	**Condition 3: Disyllabic**	**Condition 4: Monosyllabic**
doorways	deep ways	daisies	doves
dishcloths	drab cloths	dolphins	ducks
dustpans	dark pans	donkeys	drapes
graveyards	green yards	gospels	grapes
grandstands	grey stands	griddles	graphs
grindstones	good stones	goblins	gloves
nightgowns	nice gowns	noodles	nuns
neckties	neat ties	napkins	nets
nightshirts	new shirts	nickels	nodes
lampshades	low shades	lemons	leeks
logbooks	late books	lanterns	lungs
lipsticks	large sticks	lions	lanes
bookshops	big shops	barrels	bowls
bathtubs	bright tubs	blankets	blades
ballrooms	blue rooms	bankers	brooms

Word familiarity was confirmed by the same method as for Experiments 1 and 2: a judgement task was run on 36 native Southern British English speakers. This task collected rating data on all 60 experimental items as well as 12 filler items. Once again, participants were asked to rate each word on a scale modelled on the Likert-Type Scale Response Anchors (Vagias, 2006). Participants rated their familiarity on a scale from 1 to 5, where 1 represented "not at all familiar" and 5 represented "extremely familiar". Results for the judgement task can be found in Table 6.2. Word familiarity was high for all items. Crucially, the mean judgement ratings for each condition did not significantly differ from each other (p=.668). Measures of word frequency and word length were collected from the N-Watch statistical software (Davis, 2005). Frequency counts were each item's CELEX (COBUILD) frequency (the combined written and spoken counts). Further measures included the (base 10 plus 1) log frequency (also CELEX) and word length (number of letters).

Table 6.2: Weighted Averages of Word Familiarity: Experiment 3

Cond 1:	Score	**Cond 2:**	Score	**Cond 3:**	Score	**Cond 4:**	Score
doorways	5	deep ways	5	daisies	5	doves	5
dishcloths	5	drab cloths	5	dolphins	5	ducks	5
dustpans	5	dark pans	5	donkeys	5	drapes	4.97
graveyards	5	green yards	5	gospels	4.97	grapes	5
grandstands	4.97	grey stands	5	griddles	4.83	graphs	5
grindstones	4.86	good stones	5	goblins	4.97	gloves	5
nightgowns	5	nice gowns	5	noodles	5	nuns	5
neckties	4.86	neat ties	5	napkins	5	nets	5
nightshirts	4.97	new shirts	5	nickels	4.97	nodes	5
lampshades	5	low shades	5	lemons	5	leeks	5
logbooks	5	late books	5	lanterns	5	lungs	5
lipsticks	5	large sticks	5	lions	5	lanes	5
bookshops	5	big shops	5	barrels	5	bowls	5
bathtubs	5	bright tubs	5	blankets	5	blades	5
ballrooms	5	blue rooms	5	bankers	5	brooms	5
Mean Score:	4.977		5		4.982		4.998

Table 6.3: Mean Variables: Experiment 3

	Comp 1	**Comp 2**	**Comp**	**Adj**	**N**	**Di**	**Mono**
CELEX frequency	130.5	170.4	0.5	266.6	170.4	8.4	10.4
Log Frequency	1.7	1.8	0.2	2.2	1.8	0.8	0.8
No of Letters	4.1	5.1	9.3	4.1	5.1	6.2	4.4

Once again, the individual morphemes of compounds and phrases were matched as closely as possible for frequency and length: e.g. the first morpheme of the compound (e.g. *grand*) was matched for length and frequency to the adjective of the phrasal condition (*green*). To further control for any effects related to these factors, the second morpheme of the compound condition remained identical to the second morpheme of the phrasal condition (e.g. *graveyards* and *grey yards*).

6.2.3 Participants and Procedure

Twenty participants between the ages of 18 and 29 took part in Experiment 3. Participants were monolingual speakers of Southern British English. They were recruited through mailing lists at the University of Oxford and were compensated for their participation. Participants were tested individually in a soundproof booth at the University of Oxford. They were told that they would see words on the screen, then hear five different questions (see Table 6.4). The rest of the experimental design was identical to Experiment 1.

Table 6.4: Auditory Prompts: Experiments 3 and 4

1. What are good?
2. What are nice?
3. What are clean?
4. What are dry?
5. What are big?

6.2.4 Design

The experiment was designed using E-Prime 2.0 software. There were 10 blocks of 19 trials in total (190 items total including filler words) with optional breaks between the blocks. Items were distributed pseudo-randomly within the blocks, and five different versions of the experiment were presented, in which the sets of experimental questions were rotated. Each experimental word was presented once at each preparation latency:

800 ms, 1200 ms, and 1400 ms. Thus, each word appeared three times in the course of the experiment.

6.2.5 Results

6.2.5.1 Latencies Analysis

Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error. Responses uttered before the final beep were also discarded. The total number of errors was low, with only 5.3% of data being discarded due to these parameters. Data trimming treated all data points beyond two standard deviations from the mean as outliers: this resulted in the removal of an additional 129 (3.6%) data points.

The RT data were submitted to a linear mixed models analysis, in which naming latency (RT) was modelled as a function of the fixed effect factors, condition type (cond) and preparation time (beep). Subjects and items were treated as random factors. Following Bates et al. (2014), we began by testing the maximally-appropriate random structure and continued to adjust the model until it reached convergence. The maximal model contained a full interaction between condition and preparation time in the fixed effects, an interaction with random slopes and intercepts for condition and beep in the by-subjects analysis, and random intercepts for condition and beep in the by-item analysis. This model[1] failed to converge. The next maximal model to converge contained an interaction in the fixed effect factors, an interaction with random slopes and intercepts for the by-subject analysis, and random intercepts for the by-items analysis.[2] This model appeared comfortably homoscedastic when fitted to residual plots, as shown in figure 6.5. The scatterplot showed some evidence of clustering and lines; however, the histogram appeared quite well-distributed and the QQ-plot was straight. Crucially, a log-transform

[1] lmer(rt ~ cond*beep + (1+cond*beep|sub)+ (cond+beep|item)).
[2] lmer(rt ~ cond+beep + (1+cond+beep|sub)+ (1|item))

did not result in a better model fit.

Following Baayen (2008), all t-values greater than 2 were treated as significant. The results from the mixed-effect analysis of condition on reaction time revealed a significant difference in the naming latencies for adjective-noun phrases (t= 10.637*). In order to generate pairwise comparisons between the four conditions, each condition type was systematically treated as the intercept. Adjective-noun phrases differed significantly from the other three conditions, while noun-noun compounds showed no significant difference from either of the morphologically-simple word conditions. These analyses also showed that simple words differed significantly from phrases, but not from one another. These values can be seen in Table 6.5, where significant effects are marked with an asterisk. Mean reaction times and percentage error rates for Experiment 3 are shown in Table 6.6. The graph below (Figure 6.7) shows the distribution of mean naming latency (RT) for all four conditions.

The effect of preparation time (beep) was significant for the shortest preparation time (800 ms). Latencies for these items were significantly longer than for the other two preparation times (1200 ms, and 1400 ms).[3] The results also revealed an effect of block on RT: latencies were significantly faster in Block 10 than Block 1 (Estimate= -13.4730, SE= 5.3436, t= -2.521*) (Figure 6.6), however there was no interaction of condition and block on naming latencies (t= -1.38). This experiment also presented five different auditory prompts; therefore an analysis of an interaction of prompt type was carried out: none of the prompts produced an interaction with condition on RT (all ts < 1.5).

[3]In this case, the maximal model to converge included an interaction of beep and condition in the fixed effects: rt $\sim$ beep+cond+ (1+cond+beep|sub)+ (1|item)

Table 6.5: Linear Mixed-Effects Analyses of RT Data in Experiment 3

	Estimate	**SE**	**t Score**
All conditions:			
(Intercept)	344.4992	20.7910	16.570
Adj-Noun Phrases	68.0408	6.3963	10.637*
Disyllabic	-5.7598	3.5985	-1.601
Mono	0.5118	3.5956	0.142
beep2	-18.1924	5.3979	-3.370*
beep3	-20.3117	5.1094	-3.975*
Pairwise Comparisons:			
Adj-Noun Phrases as intercept:			
(Intercept)	412.539	23.760	17.363
Compounds	-68.053	6.350	-10.717*
Disyllabic	-73.790	6.956	-10.608*
Monosyllabic	-67.531	6.296	-10.725*
Disyllabic as intercept:			
(Intercept)	338.739	20.175	16.790
Compounds	1.623	6.266	0.259
Adj-Noun Phrases	42.295	14.212	2.976*
Monosyllabic	7.458	7.670	0.972
Monosyllabic as intercept:			
(Intercept)	458.213	26.687	17.170
Compounds	5.760	3.598	1.601
Adj-Noun Phrases	73.801	6.985	10.565*
Disyllabic	6.272	3.962	1.583

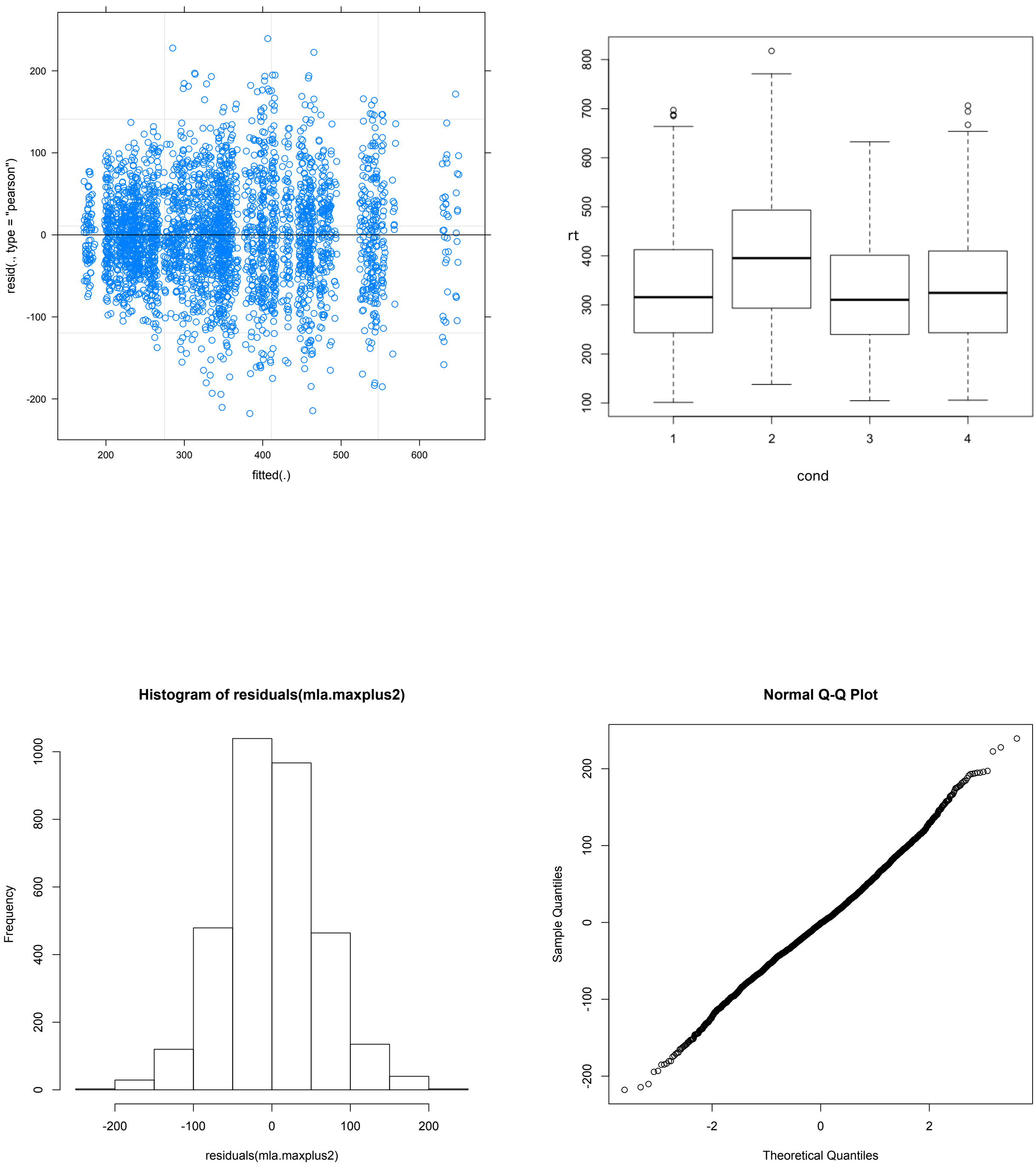


Figure 6.2: Residuals Analyses: Experiment 3

Table 6.6: Mean Naming Latencies (in ms): Experiment 3

	(1) Compounds +Clitics	(2) Phrases +Clitics	(3) Disyll +Clitics	(4) Monosyll +Clitics	Mean Lat. (ms)
PWds	1	2	1	1	
LexWs	2	2	1	1	
Syllables	3	3	3	2	
Beep Lat.:					
800 ms.	343 (7.3%)	416 (4.0%)	333 (4.7%)	344 (5.0%)	359 (4.9%)
1200 ms.	328 (5.0%)	394 (4.3%)	322 (4.3%)	325 (7.3%)	342 (5.3%)
1400 ms.	322 (6.7%)	390 (4.3%)	321 (6.3%)	327 (4.7%)	341 (5.8%)
Mean Lat. (ms)	331 (6.3%)	400 (4.2%)	326 (5.1%)	332 (5.6%)	

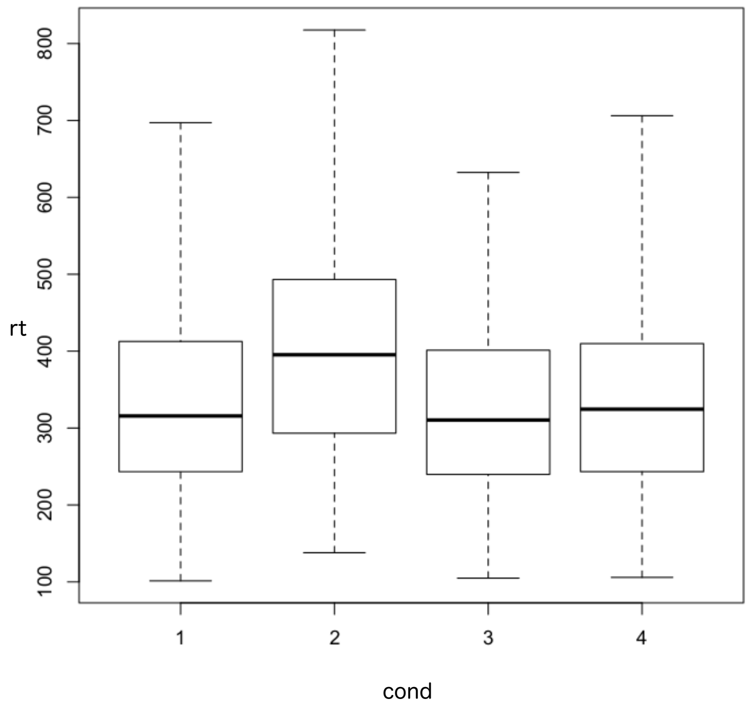


Figure 6.3: Mean RTs by Condition: Experiment 3

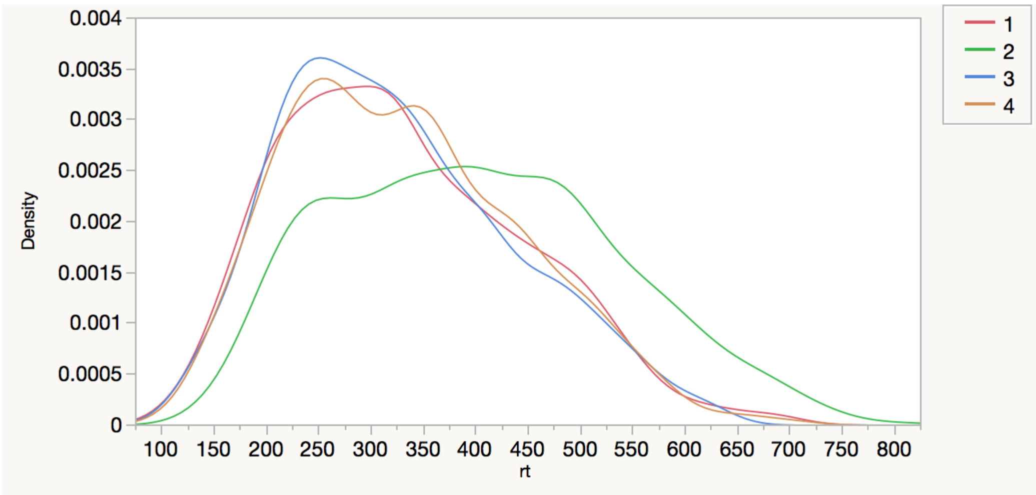


Figure 6.4: Densities of RT by Condition: Experiment 3

Table 6.7: Linear Mixed-Effects Analyses of Preparation Time*RT in Experiment 3

	Estimate	**SE**	**t score**
800 ms as Intercept			
(Intercept)	343.8921	20.9563	16.410
1200 ms	-17.2971	7.0889	-2.440*
1400 ms	-19.3863	6.8995	-2.810*
1200 ms as Intercept:			
(Intercept)	326.5950	22.9924	14.204
800 ms	17.2971	7.0889	2.440*
1400 ms	-2.0891	5.9021	-0.354
1400 ms as Intercept			
(Intercept)	324.5058	22.7067	14.291
800 ms	19.3863	6.8995	2.810*
1200 ms	2.0891	5.9021	0.354

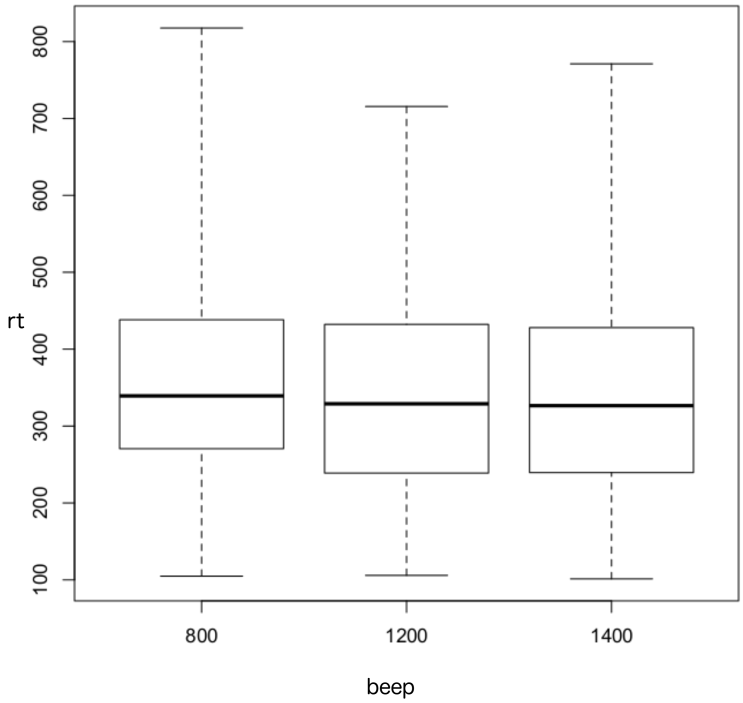


Figure 6.5: Mean RTs by Preparation Time: Experiment 3

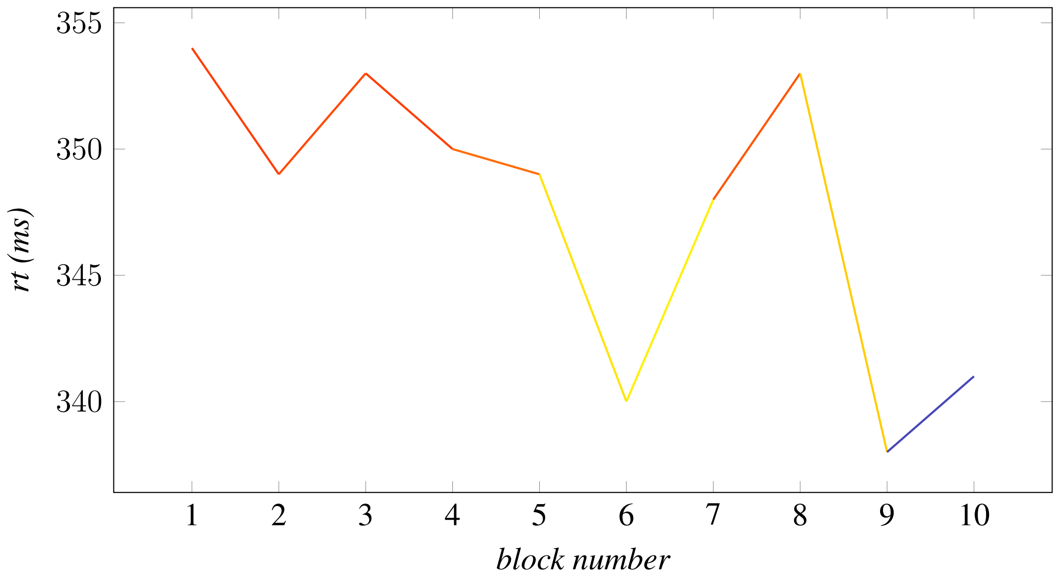


Figure 6.6: Mean RTs by Block: Experiment 3

6.2.5.2 Errors

As part of this investigation, an error analysis was carried out: errors were categorised as "time out" (the subject said nothing), "voice key error", "disfluencies" (e.g. stuttering), "stress error" (incorrect prosody), and "wrong item". While these speakers did not make frequent errors, disfluency errors were the most common, followed by wrong items and time outs (see Figure 6.7 below). The error rates were analysed proportionally using a generalised linear model (GLM) with a binomial distribution (link= logit). This analysis revealed no main effect of condition on error rate (χ^2= 2.36, p= 0.66). Nor was there a main effect of preparation time (χ^2= .04, p= .83) or block (χ^2= .16, p= .99) on error rates. While these speakers did not make frequent errors, disfluency errors were the most common, followed by wrong items and time outs (see Figure 6.8 below).

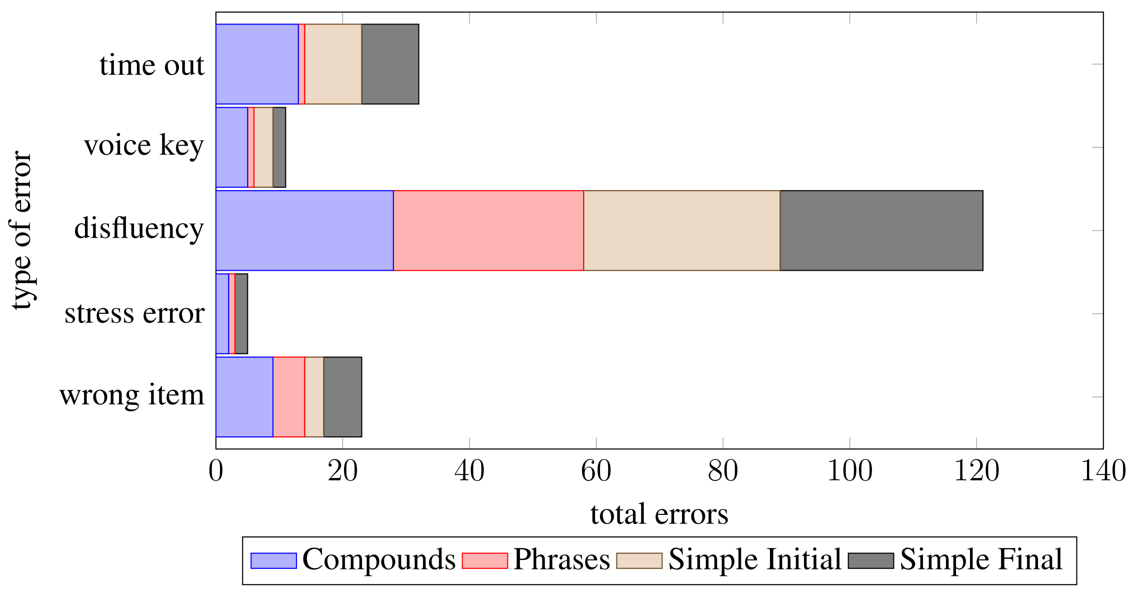


Figure 6.7: Total Errors: Experiment 3

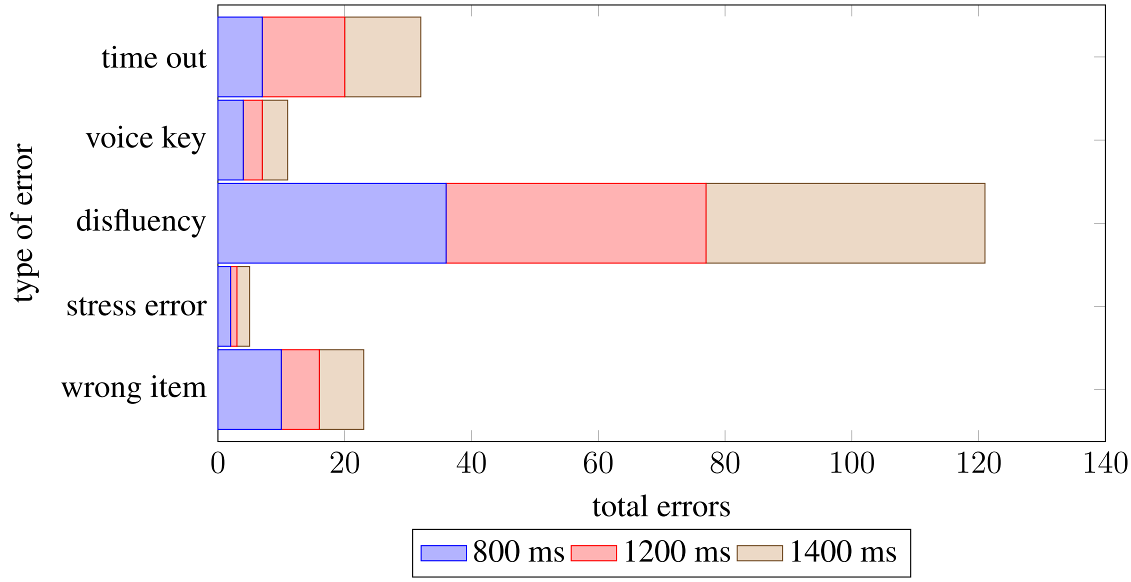


Figure 6.8: Error by Preperation Time: Experiment 3

6.2.5.3 Frequency Analyses

For this analysis, the RT data were again submitted to a linear mixed models analysis, in which frequency, condition, and preparation time (beep) were all treated as the fixed effect factors. Once again, the maximally-appropriate structure that converged presented random intercepts and slopes for the by-subject factor and random intercepts only for the by-items factor.[4] Neither log (Estimate= -0.1090, SE= 0.1645, t= -0.663) nor CELEX (Estimate= -0.2149, SE= 0.1628, t= -1.320) measures of total word frequency had an effect on naming latencies in this task. In the compound frequency analysis, neither the first (Estimate= -0.02419, SE= 0.01557, t= -1.554) nor second morpheme (Estimate = 2.497e-03, SE= 8.265e-03, t= 0.302) word frequency affected naming latency, nor did the the adjective (Estimate= 2.417, SE= 4.112, t= -0.588) or noun frequency (Estimate= -1.469, SE= 3.752 t= -0.392) of the adjective-noun phrase. There was no interaction of length and condition (Estimate= 2.406, SE= 1.834, t= 1.312). In a simple analysis of word length on all RTs (not interacting with condition), greater word length did result in longer reaction times.

[4] lmer(rt ∼ (cond+beep) + (cond+prompt) + (1+cond|sub)+ (1|item).

Table 6.8: Word Length Effect on RT: Experiment 3

Number of Letters	**Estimate**	**SE**	**t score**
(Intercept)	321.201	27.786	11.560
length5	10.824	18.689	0.579
length6	10.413	18.943	0.550
length7	6.962	18.476	0.377
length8	9.508	18.477	0.515
length9	48.473	18.945	2.559*
length10	45.517	18.471	2.464*
length11	56.615	18.690	3.029*
length12	65.668	28.788	2.281*

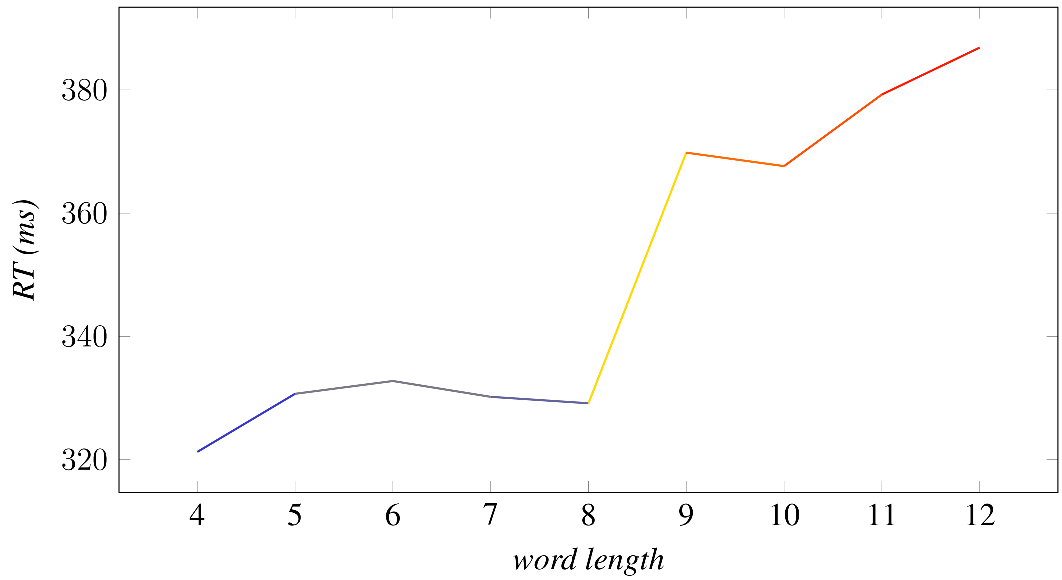


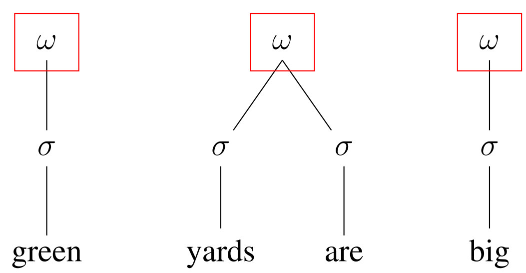
Figure 6.9: Mean RT by Word Length: Experiment 3

6.2.6 Discussion

The results from Experiment 3 showed a significant main effect of condition type on mean naming latencies. Speakers consistently took longer to produce utterances containing clitics plus adjective-noun phrases than any other condition in the task. This finding was consistent with the results from Experiment 1, which indicated that speakers were planning compounds and phrases differently. Naming latencies for the clitic + compound condition were much shorter, similar to those for both morphologically-simple word conditions. This suggests that speakers planned these three conditions similarly, at least at the phonological encoding level.

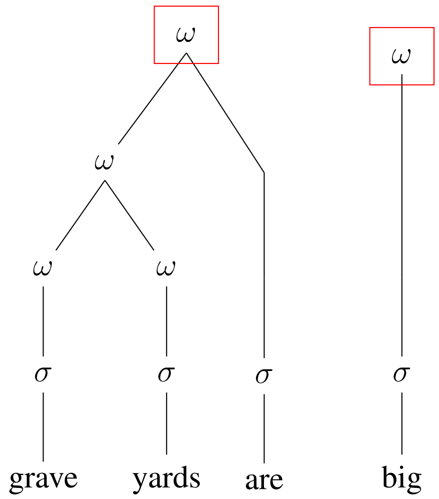
Returning to the predictions made in §6.2.1, we find evidence that, despite the increase in the complexity of the target stimuli, latencies continued to reflect the planning of prosodic units, not lexical units. For adjective noun phrases, the target items consisted of two prosodic units: the adjective, and the noun + clitic Phonological Word structure (the total utterance contains three prosodic units) (example 6.4):

(6.4) Units predicting naming latencies in adjective-noun phrases

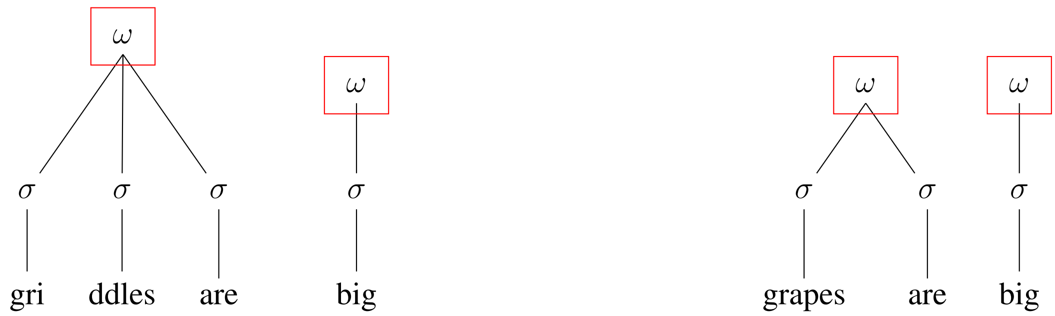


Crucially, for our purposes, we also found that the compounds behave as they did in Experiments 1 and 2: as single prosodic units. Although compounds are structurally much more complex (see Example 6.5) than either morphosyntactically-simple word conditions (Example 6.6), speakers did not treat them any differently in terms of naming latencies.

(6.5) Units predicting naming latencies in compounds



(6.6) Units predicting naming latencies in morphosyntactically-simple words



The representations in examples 6.5 and 6.6 show the phonological units (in red) that are generating the naming latencies for the compound + clitic condition, and morphosyntactically-simple word + clitic conditions. In the compound representation, it is the recursive Phonological Word unit of the compound as a whole to which the clitic attaches.

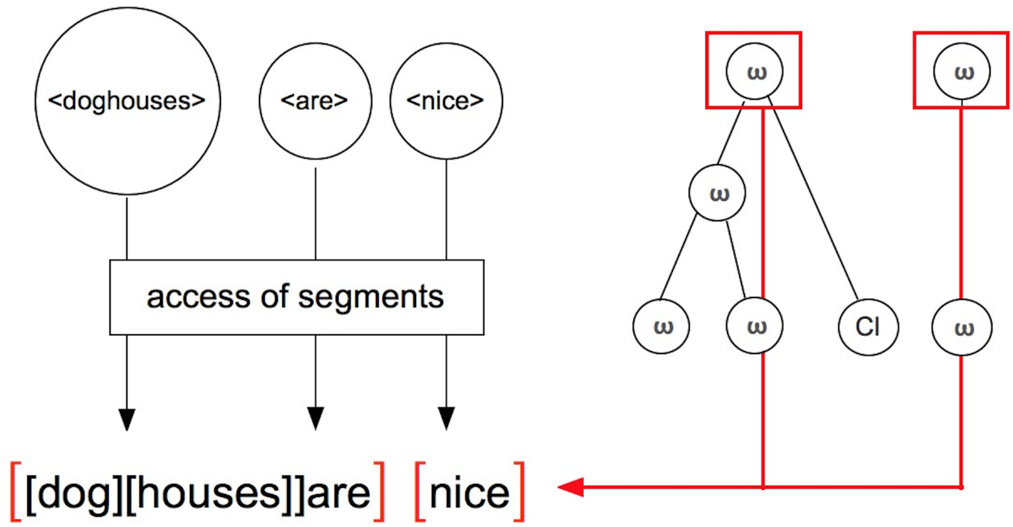


Figure 6.10: Encoding of "doghouses are nice" in WEAVER++

If we envision this process in WEAVER++, we see that the prosodic frames for *doghouses are nice* simply consist of two Phonological Word units: one for the recursive compound + clitic structure, and one for the simple word *nice*.

There was no interaction of any type between word frequency and condition type on the naming latencies. Neither the CELEX whole frequency (t= -1.320) nor log frequency (t= -0.663) produced an interaction. Nor were the measures related to constituent frequency effect significant in the compound condition: neither the left morpheme frequency (t= -1.554) nor right morpheme frequency (t= 0.302) had an effect on latencies. Word length was observed to have no interaction with condition (t= 1.312) on RT; however, as expected, word length had a general effect. Larger words generated longer naming latencies (at least for words containing 9 or more letters). In the experimental block analysis, latencies were significantly faster in block 10 than all other blocks.

Results from the error analysis were similar to those in Experiment 1: errors in this task were generally low, making up just 5.3% of the data. The most common type of error across all four conditions was "disfluency", followed by "time out", and "wrong item".

Compounds generated the highest number of errors, as did the longest preparation time (1400 ms). Reaction times were slowest for the shortest preparation time (800 ms). This bring us back to the discussion of planning scope in language production. Recall that each stage of production has associated buffers: phonological encoding has an input and output buffer where unfolding phonological segments are compiled and stored. These buffers have a limited capacity due to the constraints of working memory, and their content becomes degraded as time passes (cf. Wagner et al., 2010).

The increase in complexity of both task and target items was evident from the results of Experiment 3: we saw an effect of the shortest preparation time, as well as an effect of word length on naming latencies. The difference between mean naming latency for compounds and phrases was evident: utterances containing phrases and clitics took on average 70 ms longer to prepare. However, as with the results of Experiment 1, this data only tells us what happens when speakers have time to prepare their utterances in advance. Specifically, it does not reveal anything about the direction of clitic attachment: the total number of prosodic units predicting the naming latencies remains the same irregardless of which unit the clitic is attaching to:

(5.1) **Left clitic attachment:** (doorways are)$_\omega$ (clean)$_\omega$
Right clitic attachment: (doorways)$_\omega$ (are clean)$_\omega$

In English, the default direction of auxiliary attachment is leftwards (cf. Labov, 1969; Zwicky, 1977); however, we must generate concrete evidence for this, particularly in regards to the compound conditions, if we are to use this task to test planning in L2 speakers. We saw the effect of the removal of planning time on the prosodic planning in Experiment 2: speakers' naming latencies reflected only the first prosodic word of the utterance and, furthermore, the size of this word predicted the latencies. Our next experiment seeks to elicit the same effect on these new target items.

6.3 Experiment 4: Online Production of Clitics

6.3.1 Hypotheses

The results from Experiment 3 indicate that clitics are attaching to the larger, recursive prosodic structure of compounds. This suggests that the auxiliary verb is behaving as a clitic, becoming unstressed and cliticising leftwards onto prosodic units: for compounds, the unit of planning was a recursive higher Phonological Word constituent. Following these results as well as keeping in mind the results of the online task in Experiment 2, we propose the following predictions below:

1. The loss of planning time will affect the size of the planning unit as well as the process that prepares it for articulation.

We saw the effects of restricted planning time in Experiment 2, where speakers generally made more errors and exhibited longer naming latencies over the course of the experiment. We also saw the effects on the construction of prosodic frames: speakers were only able to plan the first prosodic frame before the utterance began. Crucially, however, they would sacrifice initiation time to ensure that the first frame contained not a lexical word, but a well-formed Phonological Word.

Here, we predict that speakers will do the same: they will continue to plan speech from left to right, in frames that are organised according to prosodic, not lexical word structure. Because speakers will only be able to encode the first prosodic frame, the latencies for adjective-noun phrases will be significantly shorter than for all other conditions.

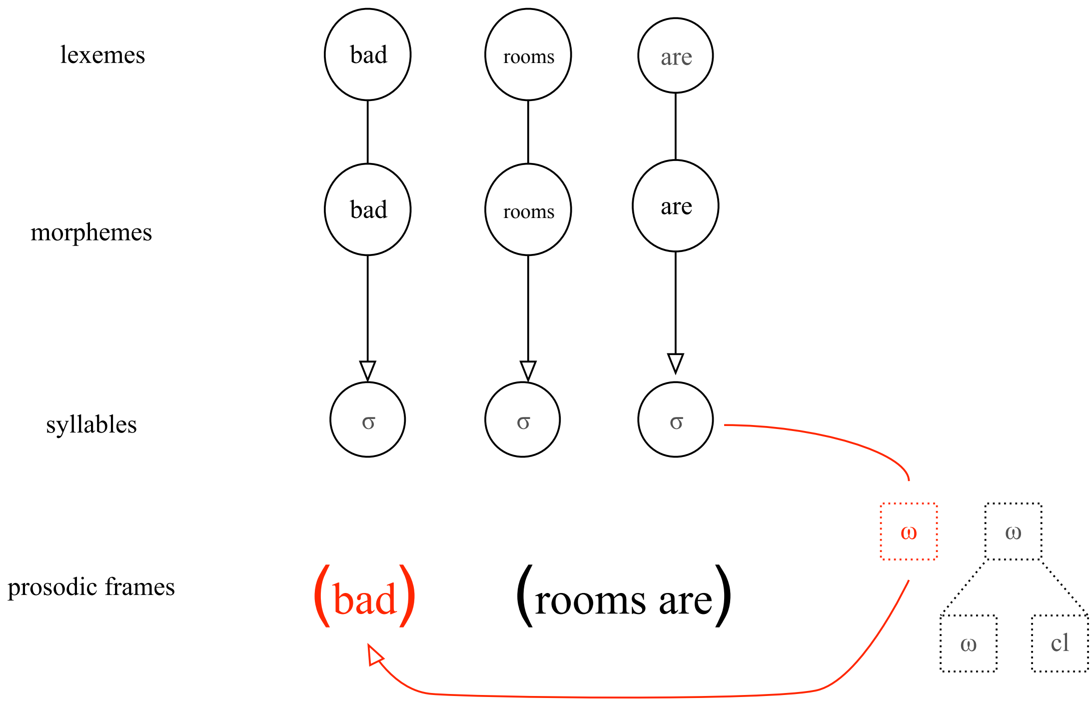


Figure 6.11: Encoding of "bad rooms are" in WEAVER++

Experiment 3 indicated that clitics only attach to the noun in the phrasal condition; therefore, as the Formulator prepares its prosodic frames from left to right, it only needs to access the adjective in this utterance. This will result in significantly shorter times than all other conditions, which will include clitics in their Phonological Word representations.

2. **Compounds will retain their recursive word shape, and continue to elicit naming latencies similar to simple morphosyntactic words.**

Experiment 3 indicated that the compound + clitic condition was being prepared as a single prosodic unit, containing several layers of recursion. In this task, we predict that speakers will continue to plan these items as single prosodic units, resulting in latencies that reflect a more complex phonological unit than is found in the phrasal condition. However, the complexity of this unit will not cause naming latencies for compounds to differ from the morphosyntactically-simple words. Because all three conditions will involve (Phonological Word+ auxiliary), they will elicit similar naming latencies.

3. The auxiliary will reduce and attach leftwards onto the Phonological Word unit. This hypothesis is based on a great deal of existing evidence for auxiliary reduction in English, namely that this phenomenon is more likely to happen in certain environments. Selkirk writes,

> "The phonetic form of the contraction... is determined by exactly the same principles that govern the phonetic form of the plural, the verb inflection, and the possessive... i.e..., by what the final segment of the preceding noun (or element of N") or verb is (1972: 79)

We also know that certain syntactic environments block auxiliary reduction (cf. the discussion in §2.3.1). As such, we hypothesise that the auxiliary verb *are* will reduce and attach leftwards to prosodic units in this experiment. Since the naming latencies will only reflect the size of the first prosodic unit, we will be able to confirm the direction of cliticisation in addition to cliticisation of the auxiliary itself.

6.3.2 Stimuli Selection and Design

Experiment 4 presents the target words of Experiment 3 in online task conditions. The experiment was designed using E-Prime 2.0 software. There were 6 blocks of 12 trials (72 items total including 12 filler words) with optional breaks between the blocks. Items were distributed pseudo-randomly within the blocks, and five different versions of the experiment were presented in which the sets of experimental questions were rotated. As in Experiment 2, the experimental stimuli were only presented once in this task.

6.3.3 Participants

Twenty-five participants between the ages of 18 and 26 took part in Experiment 4. Participants were monolingual speakers of Southern British English. They were recruited through mailing lists at the University of Oxford and were compensated for their participation. They were tested individually in a soundproof booth at the University of Oxford.

6.3.4 Results

6.3.4.1 Latencies Analysis

Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error. Utterances which did not exhibit reduction and cliticisation were also marked as errors: this was done by measuring the duration of the vowel in the auxiliary verb. Responses uttered before the final beep were also discarded. The total number of errors was very low, with only 3.8% of data being discarded due to these parameters. Data trimming treated all data points beyond two standard deviations from the mean as outliers: this resulted in the removal of an additional 56 data points. The total loss made up only 7.8% of data.

The RT data were submitted to a linear mixed models analysis, in which naming latency (RT) was modelled as a function of the fixed effect factors and condition type (cond). Subjects and items were treated as random factors. Following Bates et al. (2014), we began by testing the maximally-appropriate random structure and continued to adjust the model until it reached convergence. The maximal model contained condition in the fixed effects, random slopes and intercepts in the by-subject analysis, and intercepts for condition in the by-item analysis. The next maximal model to converge contained condition as a fixed effect, random slopes and intercepts for subjects, and intercepts only for the by-item analysis.[5] This model's values can be found in Table 6.9, where significant effects are marked with an asterisk. This model appeared comfortably homoscedastic in the residuals histogram plot, shown in Figure 6.12. A log-transform did not improve fit.

Following Baayen (2008), all t-values greater than 2 were treated as significant. The results from the mixed-effect analysis of condition on reaction time revealed a significant difference in the naming latencies for adjective-noun phrases (t= -2.403*) and a significant difference in monosyllabic words + clitics (t= -2.410*). In order to generate pairwise comparisons between the four conditions, each condition type was systematically treated as the intercept. Adjective-noun phrases differed significantly from the other three conditions, while noun-noun compounds showed no significant difference from either of the morphologically-simple word conditions. These analyses also indicated that simple words differed significantly from phrases, but not from one another. Mean reaction times and percentage error rates for Experiment 4 are shown in Table 6.10. The graph below the table (Figure 6.13) shows the distribution of mean naming latency (RT) for all four conditions. There was no interaction between block and condition on naming latencies; however, latencies increased significantly in the final three blocks (cf. Table 6.11). There was no evidence (t= 0.128) of an interaction between prompt type and condition; however, naming latencies were significantly faster overall (t= -2.696*) for the prompt, "What are nice?" (see Figure 6.15).

[5] mla.maxplus3 <- lmer(rt ~ cond + (1+cond|sub)+ (1|item))

Table 6.9: Linear Mixed-Effects Analyses of RT Data in Experiment 4

	Estimate	**SE**	**t value**
All Conditions:			
(Intercept)	229.161	18.050	12.696
Adj-N Phrases	-31.375	13.057	-2.403*
Disyllabic	-12.827	12.135	-1.057
Monosyllabic	-26.824	11.133	-2.410*
Adj-N Phrases as Intercept:			
(Intercept)	192.691	19.094	10.092
Compounds	25.825	8.890	2.905*
Disyllabic	30.095	8.253	3.647*
Monosyllabic	16.560	7.458	2.221*
Disyllabic Simple Words as Intercept:			
(Intercept)	222.786	18.082	12.321
Compounds	-4.269	7.258	-0.588
Adj-N Phrases	-30.095	8.253	-3.647*
Monosyllabic	-13.535	6.277	-2.156*
Monosyllabic Simple Words as Intercept:			
(Intercept)	209.251	17.454	11.988
Compounds	22.348	11.1325	2.410*
Adj-N phrases	-16.560	7.458	-2.221*
Disyllabic	13.535	6.277	2.156*

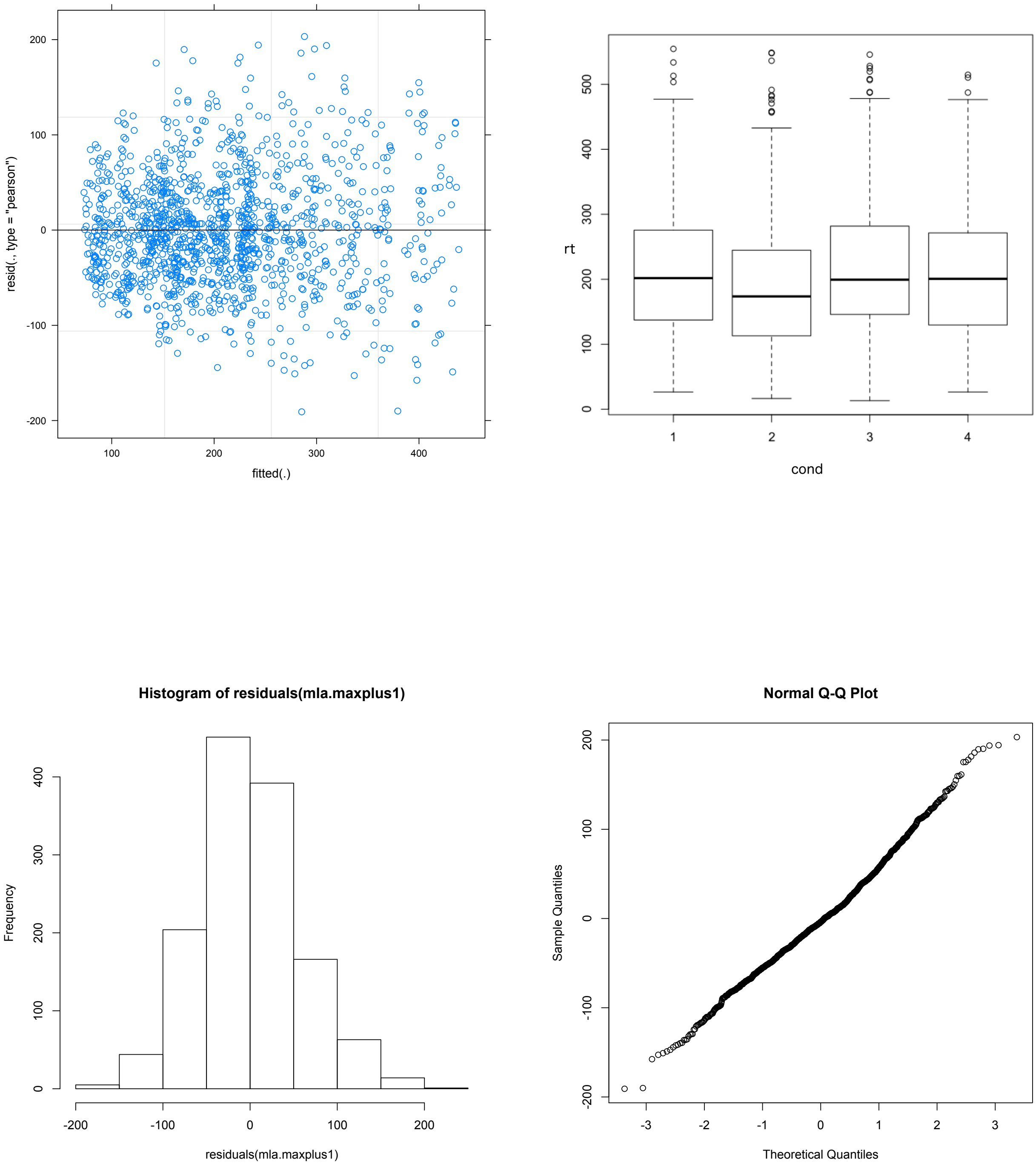


Figure 6.12: Residuals Analyses: Experiment 4

Table 6.10: Mean Naming Latencies: Experiment 4

	(1) Comp +Clitics	**(2) Phrases +Clitics**	**(3) Disyll +Clitics**	**(4) Mono +Clitics**
PWds	1	2	1	1
LexWds	2	2	1	1
Syllables	2	1	2	1
Mean. Lat. (ms)	215 (4.2%)	190 (3.2%)	222 (4.0%)	208 (4.3%)

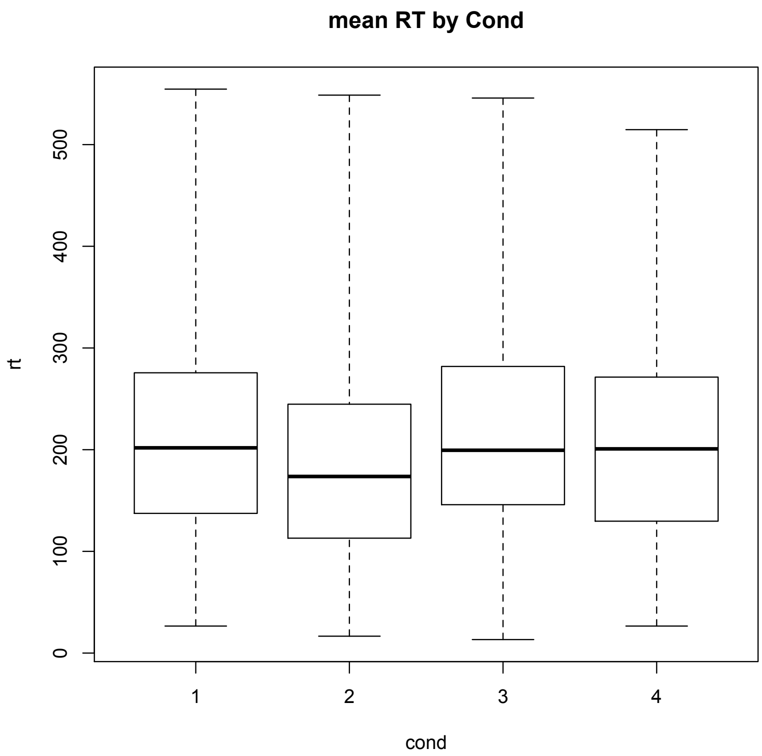


Figure 6.13: Mean RTs by Condition: Experiment 4

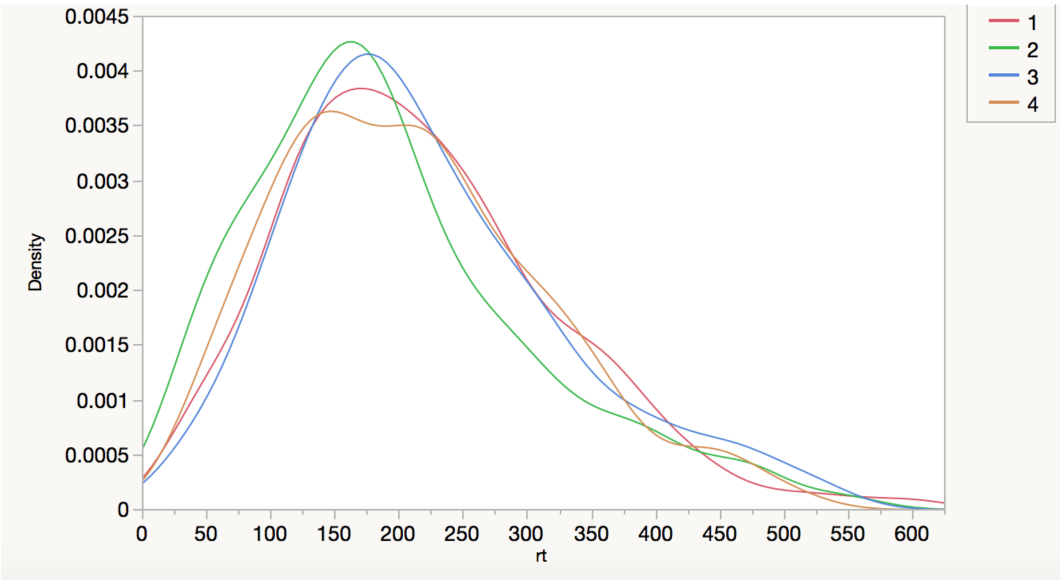


Figure 6.14: Densities of RTs by Condition: Experiment 4

Table 6.11: Linear Mixed-Effects Analyses of Mean RT by Block: Experiment 4

	Estimate	**SE**	**t score**
(Intercept)	215.156	13.508	15.928
block2	4.385	4.420	0.992
block3	9.496	4.366	2.175*
block4	14.277	4.418	3.231*
block5	11.166	4.388	2.545*
block6	18.224	4.300	4.239*

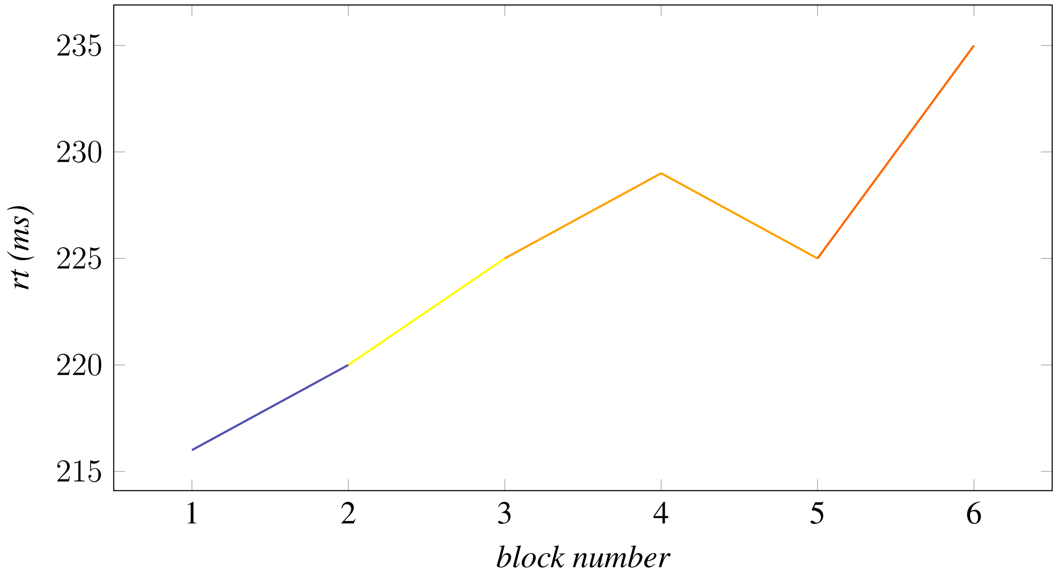


Figure 6.15: Mean RT by Block: Experiment 4

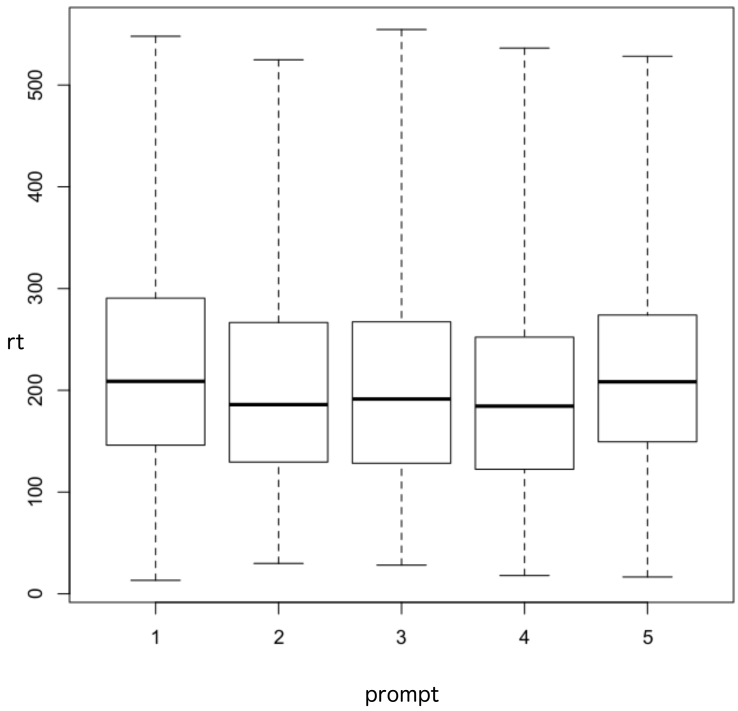


Figure 6.16: Mean RTs by Prompt: Experiment 4

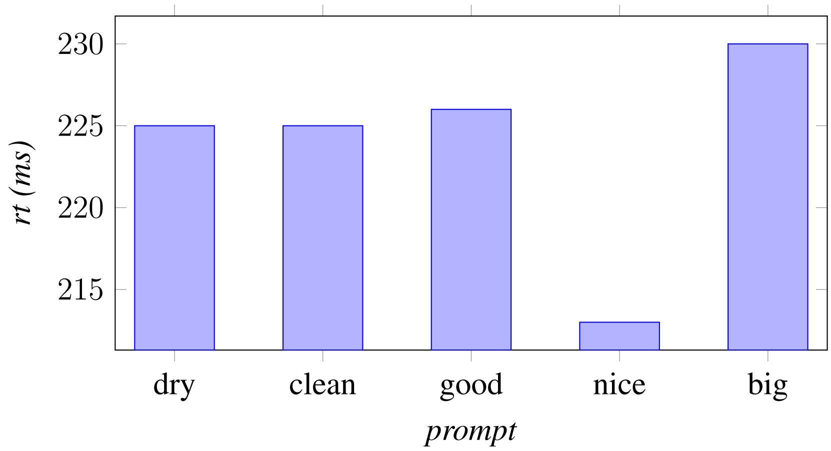


Figure 6.17: Mean RTs by Prompt: Experiment 4

6.3.4.2 Errors

As part of this investigation, an error analysis was carried out: errors were categorised as "time out" (the subject said nothing), "voice key error", "disfluencies" (e.g. stuttering), "stress error" (incorrect prosody), and "wrong item". The error rates were analysed proportionally using a generalised linear model (GLM) with a binomial distribution (link= logit). There was no main effect of condition on error rates (χ^2= 5.77, p= <.12). Disfluencies also made up the largest portion of errors here. While it appeared that blocks 5 and 6 generated the highest number of errors of all blocks, there was also no main effect of block on error rates (χ^2= 2.87, p= <.09).

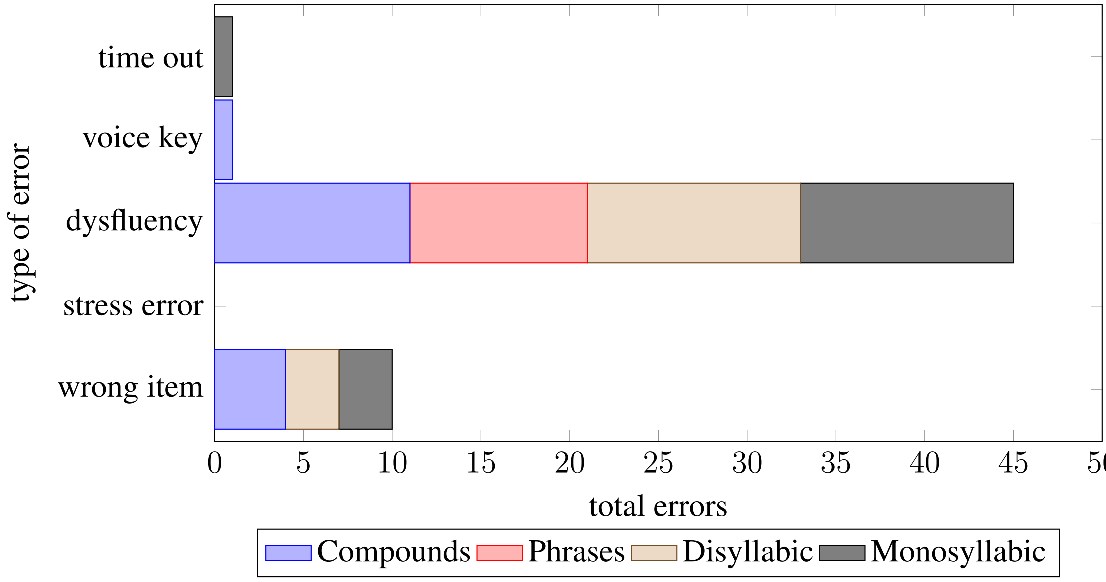


Figure 6.18: Total Errors: Experiment 4

6.3.4.3 Frequency Analyses

For this analysis, the RT data were again submitted to a linear mixed models analysis, in which frequency and condition were treated as fixed effect factors. Once again, the maximally-appropriate structure that converged presented random intercepts and slopes for subject factors and a random intercept only for the by-items factor.[6] The models converged and their results are shown in Table 6.12 below. Neither log (Estimate= 4.220, SE= 3.662, t= 1.152) nor CELEX (Estimate= 0.0414, SE= 0.1794, t= 0.231) measures of total word frequency had an effect on naming latencies in this task. In the compound frequency analysis, neither the first (Estimate= -0.03007, SE= 0.01790, t= -1.68) nor second (Estimate= 8.100e-03, SE= 9.342e-0, t= 0.867) morpheme word frequency affected naming latency. In addition, neither the adjective (Estimate= -0.003635, SE= 0.008470, t= -0.429) nor the noun morpheme frequencies (Estimate= 4.140e-04, SE= 8.501e-03, t= 0.049) showed an interaction effect. Word length showed a strong effect on reaction times in general (cf. Table 6.12 and Figure 6.19) in that larger words took longer to plan, but it did not reveal any interaction with condition.

[6]Formula: rt $\sim$ cond + freq (1 + cond | sub) + (1 | item).

Table 6.12: Word Length and RT Analyses: Experiment 4

	Estimate	**SE**	**t score**
(Intercept)	271.467	19.668	13.802
length5	11.215	9.608	1.167
length6	8.944	9.765	0.916
length7	-5.100	9.962	-0.512
length8	-14.545	11.367	-1.280
length9	-29.788	12.090	-2.464*
length10	-33.624	11.948	-2.814*
length11	-36.702	11.915	-3.080*
length12	-48.957	15.976	-3.064*

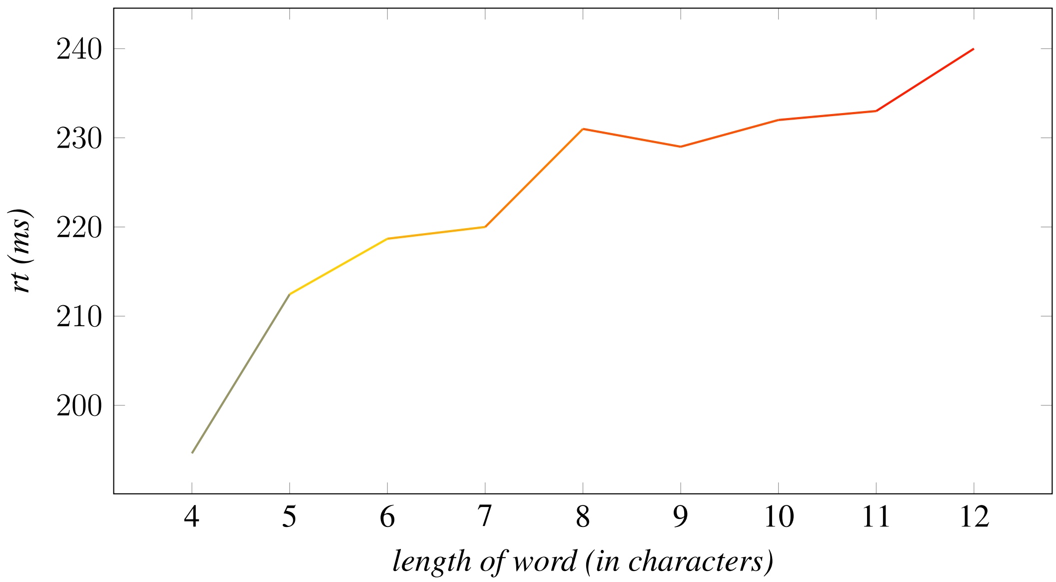


Figure 6.19: Mean RT by Word Length: Experiment 4

6.3.5 Discussion

Returning to our hypotheses for this task, we find evidence of the effect of planning scope on production: once again, speakers were only able to plan the first prosodic unit in the utterance before initiating speech. This lends support both to the incremental planning of speech (as in WEAVER++) and the shape of the frames into which segments are placed during this process. Speakers again sacrificed initiation speech to produce a well-formed Phonological Word, regularly taking more time to produce the cliticised conditions containing compounds and morphosyntactically simple words (both disyllabic and monosyllabic) than they did to produce adjective-noun phrases. In fact, adjective-noun phrases elicited a mean naming latency that was on average 30 ms faster than all other conditions.

If we return to the findings of Experiment 2, where speakers were planning the four conditions without clitics, we see some similarities. The naming latencies of the compounds and phrases were, again, significantly different (t= 5.731*) in this task, supporting our claim that these two structures were planned differently.

However, a cross-analysis of the naming latencies for compound + clitic condition **only** showed a statistically-significant result for the disyllabic control condition. The mean naming latency for the monosyllabic word + clitic condition (208 ms) was significantly faster than either the disyllabic (215 ms) or compound condition (215 ms). Adjective-noun phrases were faster yet, with a mean naming latency of 190 ms. These findings differ to those in Experiment 2, where the mean latency for the monosyllabic word condition (223 ms) was similar to that of the adjective-noun phrase (221 ms).

In the results for Experiment 2 (cf .§5.3.7), we saw that the naming latencies reflected the complexity of the first Phonological Word; therefore the mean latency for the adjective-noun phrase was nearly identical to that of the monosyllabic control condition, and the disyllabic simple word to that of the compound condition. In this task, the first prosodic unit in the utterance had three sizes:

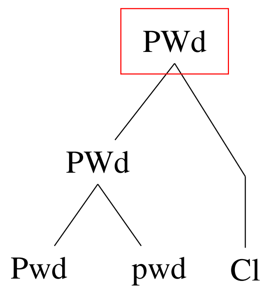
Table 6.13: Size of the Prosodic Unit in Experiment 4

Cond	First PWd Unit	Size of First PWd Unit
Compound + Clitic	(doorways are)$_\omega$	(σ σ σ)
Disyllabic + Clitic	(donkeys are)$_\omega$	(σ σ σ)
Monosyllabic + Clitic	(dates are)$_\omega$	(σ σ)
Adj-N + Clitic	(dark)$_\omega$	(σ)

Therefore the results in this experiment agree with our original hypothesis, that speakers in online production tasks are only able to access and plan the first phonological unit before initiating utterance. In this task, the condition with the smallest Phonological Word (the phrasal condition) generated the fastest response times, followed by the monosyllable + clitic condition, and then the two larger compound + clitic and disyllable + clitic structures.

Our second hypothesis was that compounds would retain their recursive word shape. Although the compound + clitic target items contained the same number of syllables and lexical words as the phrasal+ clitic items, they once again elicited distinctly different naming latencies.This bolsters our initial prediction from Experiment 2, that speakers are treating phrases as as two discrete phonological units, and only planning the first Phonological Word before initiating speech. Moreover, the similarity in naming latencies of compound words to the morphologically-simple words suggests that speakers are still planning compounds as single prosodic structures, and treating the compound + clitic structure as a single prosodic unit, as in example 6.7:

(6.7) Compound+clitic structure



Our final hypothesis concerned the direction of attachment of the auxiliary verb *are*. Empirical evidence for the reduction and leftwards attachment of this auxiliary is strong, however we wanted to ensure that speakers were exhibiting this phenomenon all four conditions. While Experiment 3 was able to tell us the total number of Phonological Words in the utterance, only Experiment 4 was able to confirm the direction of attachment. Table 6.14 presents the structure of prosodic units for both the leftwards and rightwards attachment of *are*":

Table 6.14: Size of the Prosodic Unit in Experiment 4

Cond	**Leftwards Attachment**	**Rightwards attachment**
Compound + Clitic	(doorways are) $\sigma\ \sigma\ \sigma$	(doorways) $\sigma\ \sigma$
Adj-N Phrase + Clitic	(dark) σ	(dark) σ
Disyllabic + Clitic	(donkeys are) $\sigma\ \sigma\ \sigma$	(donkeys) $\sigma\ \sigma$
Monosyllabic + Clitic	(dates are) $\sigma\ \sigma$	(dates) σ

Here, the monosyllabic word condition is key: if the auxiliary reduces and attaches leftwards, then the monosyllabic word condition should generate significantly-longer naming latencies than the phrasal condition (but shorter than the compounds or disyllabic words). If the auxiliary does not reduce, or reduces rightwards, then the monosyllabic condition would generate similar naming latencies to the phrasal condition. Accordingly, as discussed in this section above, we found evidence in the naming latencies of leftwards attachment: the monosyllabic+ clitic condition generated naming latencies that were on average 18 ms longer than the phrasal conditions. Results for the analysis of experimental block on naming latencies were wholly expected, considering the increase in working memory load that online tasks require. Naming latencies rose significantly in the last 3 blocks, evidence of the load that this task puts on reaction time. The faster response times associated with the auditory prompt, "what are nice?" are interesting. This may have been due to the difficulty of picking up the boundary related with this prompt during analysis, as naming latencies were measured from the final sound of the question prompt to the first sound of the response. However, although "dry" had a similarly unclear boundary (*What are dry?*), it did not produce the same effect. Word frequency of the prompts did not explain this either, as "good" had a significantly higher frequency than all other prompts.

Errors were surprisingly low in this task, making up just 3.8% of the data. The most common type of error across all four conditions was "disfluency", followed by "wrong item". Compounds and monosyllabic words generated the most errors- results similar to Experiment 3. There was no interaction of any type between word frequency and condition type on the naming latencies. Neither the CELEX whole frequency (t= 0.231) nor log frequency (t= 1.152) produced an interaction. Nor were the measures related to constituent frequency effect significant in the compound condition: neither the left morpheme frequency (t= -1.68) nor right morpheme frequency (t=0.867) had an effect on latencies. Remarkably, frequencies had no effect on naming latencies overall in the task; something that is contrary to many online task results.

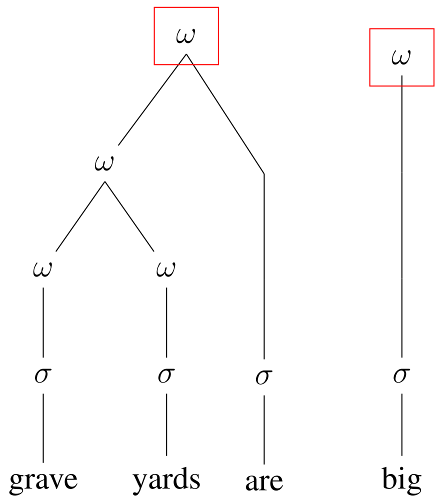
Word length had a very strong effect on naming latencies but there was no interaction between condition and length in this experiment. In comparison to Experiment 2, the task type and conditions are much more complex: participants in Experiment 2 were essentially completing a reading task. They saw the experimental item, heard the auditory prompt, and then responded right away. This task, however, involves a much longer response, five alternating auditory prompts, and more complex target items (target + clitic).

6.4 General Discussion

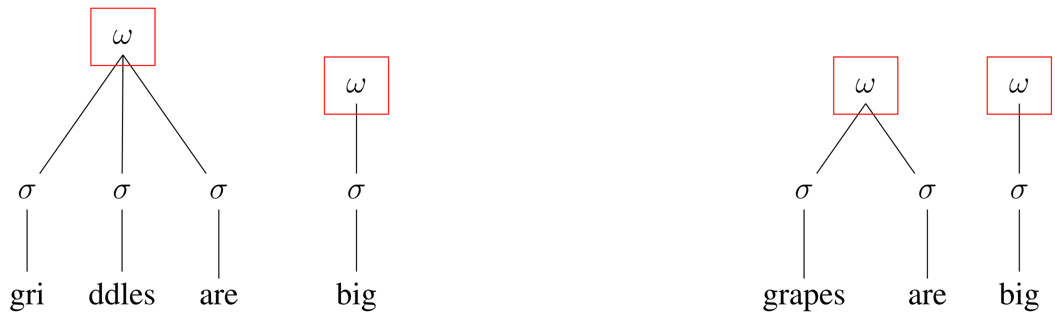
The experiments in this chapter were designed to test our Phonological Word hypothesis further. In order to do this, we introduced clitics into the utterances in Experiments 3 and 4. Using a new set of experimental stimuli (as well as a new auditory prompt), we endeavoured to create a productive environment for the reduction and cliticisation of the auxiliary verb *are*. Specifically, we wanted to see how this auxiliary would attach to the compound word stimuli: whether speakers would continue to treat compounds as single recursive prosodic units to which the auxiliary could attach, or whether the auxiliary might only attach to one element of the compound. Empirical evidence for the behaviour of both clitics and compounds suggests that compounds will retain their prosodic shape, and that clitics will reduce and attach leftwards. However, no experimental evidence exists for the production of both compounds and clitics in English: therefore, these experiments sought to address this deficit.

The results of Experiments 3 and 4 were consistent with those in Experiments 1 and 2: in the delayed task, utterances containing cliticised compounds elicited similar naming latencies to utterances containing morphologically-simple words. This suggested that speakers were planning all three utterances similarly, as in Examples 6.8 and 6.9 below.

(6.8) Units predicting naming latencies in compounds: Experiment 3

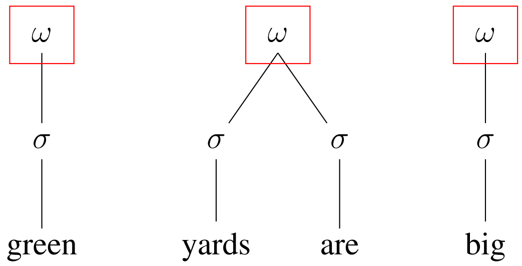


(6.9) Units predicting naming latencies in simple words: Experiment 3



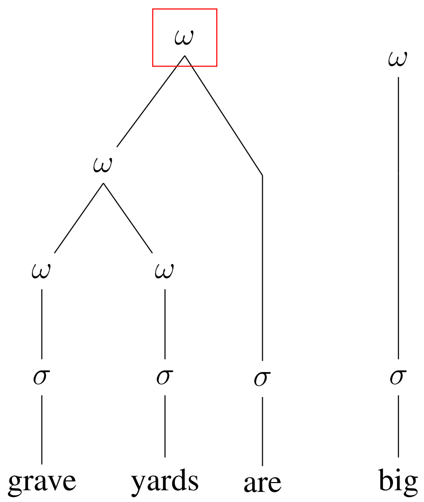
In the delayed task, adjective-noun phrases elicited significantly-longer latencies, resulting in an average 70 ms difference between compounds and phrases. This suggested that speakers planned these structures differently, as in Example 6.10:

(6.10) Units predicting naming latencies in adjective-noun phrases

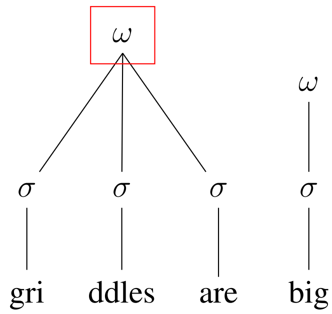


Under online task conditions (Experiment 4), speakers once again generated similar naming latencies for utterances containing cliticised compounds to those containing cliticised disyllabic words.

(6.11) Units predicting naming latencies in compounds: Experiment 4

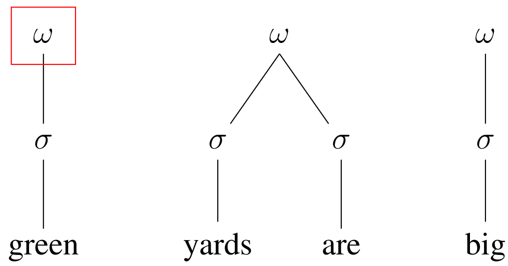


(6.12) Units predicting naming latencies in simple words: Experiment 4



Latencies for the adjective-noun phrases were significantly shorter (30 ms), indicating that speakers were only encoding the first prosodic unit of the phrase: the monosyllabic adjective.

(6.13) Units predicting naming latencies in adjective-noun phrases: Experiment 4



Taken together, the results of these two experiments indicate that compound words, despite containing two lexical words and two Phonological Words, act like a single recursive prosodic unit when speakers are encoding them with phonological features during speech production in English. Both experiments indicated that compounds were being planned as a single unit, generating significantly-different mean latencies to the adjective-noun phrases. Critically, speakers in the online tasks continued to treat compounds as single, disyllabic units, suggesting that these items retained their prosodic structure even when planning time was taken away.

As we conclude the discussion of this second set of English experiments, we will take a brief pause to look at the larger picture. The results of all four experiments establish the phonological planning unit in English; that is, we now see that native English speakers are building prosodic frames for spoken language according to prosodic, not lexical, units. Not only do we have evidence for this process in delayed naming conditions, but most importantly, we have seen how it works in online speech production. When subjected to both task and structural complexity, the planning unit remains the same: the Phonological Word. The evidence from the compound naming latencies supports current theoretical approaches to the post-lexical shape of compounds: when produced in isolation, they elicit naming latencies identical to those for both monosyllabic and disyllabic simple words. When produced with auxiliaries, the auxiliaries behave exactly as they do for morphosyntactically-simple words. Most importantly, in all four experiments, adjective-noun phrases and noun-noun compounds elicited significantly different naming latencies. We conclude this section with an answer to our first research question: from the data contained in these studies, we have solid evidence that the planning unit in English during phonological encoding is the recursive Phonological Word.

7 The Phonological Encoding of Compounds by Non-Native English Speakers

7.1 Introduction

Experiments 1, 2, 3, and 4 were designed to examine the generation of prosodic frames in native speakers of English. By comparing compounds and phrasal structures, we were able to see how speakers overwhelmingly treated compounds as single prosodic items, no matter what the task conditions or complexity. In the online tasks, they even sacrificed utterance initiation speed to prepare a well-formed prosodic structure: in the case of both compounds and compounds + clitics, this structure was always a single recursive prosodic unit. Now that we have established how prosodic frames are assigned in English, we turn to our second research question: what is the planning unit in non-native speakers of English?

The difficulty associated with producing speech in one's non-native language has long been established: for example, we know that picture-naming is often slower in a speaker's L2 than in their L1 (Kroll et al., 2006) and that error rates also increase when speaking in an L2 (Christoffels et al., 2006). In our review of L2 production models in §3.3, we saw that much of the attention paid thus far has been to the lexicalisation stage. Aside from what is discussed in the WEAVER++ L2 (2007) model, there is little mentioned about the phonological encoding processes in non-native speakers.[1]

The primary aim of the following four experiments is to isolate the prosodification process and identify the unit generated during non-native English phonological encoding. Grosjean et al. (2004) advise, in studies on non-native language to control for as many linguistic factors as possible, to reuse stimuli that have already elicited successful results, and to replicate the results using a new set of stimuli. In the selection of the stimuli for the native speakers in Experiments 1 and 2, we have already controlled for many linguistic factors (frequency, familiarity, word-length); Experiments 5 and 6 seek to test these stimuli on non-native speakers of English. By keeping the original paradigm intact, we aim to elicit data related specifically to the construction of prosodic frames and representations in L2 speech production.

We have chosen to use native speakers of Standard Colloquial Bengali as participants in both these groups. Bengali, like most languages that have been in close contact with English, has a large loanword vocabulary. However, our interest is not in how English loanwords have been incorporated into Bengali, but how complex English prosodic structures are dealt with in speakers whose native language is Bengali. Little is known about the preparation of prosodic units in non-native language. This is unfortunate, because it is at this stage that some of the most distinctive features of non-native speech are generated (e.g. syllabification and stress assignment). The experiments in this chapter seek to examine the prosodification of complex morphosyntactic structures in non-native English speakers and the questions we ask here are thus: to what extent is the preparation of the L2 prosodic structure affected by the L1 phonological processing? Are phonological encoding patterns tied to fluency?

[1]"Phonology is to a large extent the Cinderella of bilingual studies... for the non-specialist, phonology is so predominantly concerned with the mechanics of language realization, so far removed from the broader psychological issues of cognitive development and processing, that it fails to grab the attention (even of the linguist who is a non-phonologist) in the same way as matters semantic, pragmatic, or even syntactic." (Watson, 1991: 25)

The organisation of this chapter will be slightly different than that of the previous chapters. For reasons explained in §7.1.1 below, we will be testing two groups of Bengali speakers in the delayed experiment (Experiment 5). Therefore, §7.2 will contain the results and discussion of two delayed production tasks: we refer to these tasks, respectively, as Experiment 5a and Experiment 5b. Experiment 5a presents data from a highly-fluent group of L2 English speakers, while Experiment 5b presents data from a group with low fluency. Everything else in the task design remains the same. As in previous chapters, the second section (§7.3) presents results from an online experiment, and (§7.4) contains a general discussion of all three experiments.

7.1.1 Hypotheses

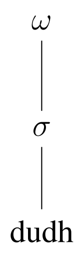
In light of previous research, the main hypotheses tested in this study are as follows:

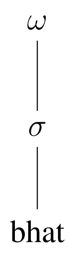
1. **Native Bengali speakers will also plan English compounds as single prosodic units.**

This hypothesis relies both on the level of fluency in English of the Bengali speakers and on their familiarity with the default prosodic structure of compounds (which we argue to be a single prosodic unit). The results of Experiment 1 indicated that, when they had time to prepare their utterances, native speakers planned compound words as simple prosodic units, no different from simple morphosyntactic words. Because speakers had ample time to plan their replies, the naming latencies for this task indicated the total number of Phonological Words in the utterance. Crucially, neither the number of lexical units nor syllables had any effect on naming latencies. This lends support to the claim that speakers plan speech in prosodic units, at least during the phonological encoding stage.

We hypothesise that Bengali speakers, if they are fluent enough in English, will have no difficulty replicating these results. First of all, the similarities between English and Bengali compound structure facilitate "structural overlap", an effect seen when shared linguistic structures (such as those in compound formation) induce cross-linguistic influence between languages (cf. Foroodi-Nejad & Paradis, 2009). Simply put, targets which contain a structure present in both L1 and L2 will facilitate naming. Children who acquire a language with very productive compounding (e.g. German and English) exhibit significantly earlier production and identification of compounds in speech than those with a less compound-productive language (e.g. Hebrew) (Clark & Berman, 1989; Nicoladis, 2002).

Compounds composed of two or more noun stems are common in Bengali. Like English compounds, the Bengali compound is composed of individual units that have their own stress when isolated:

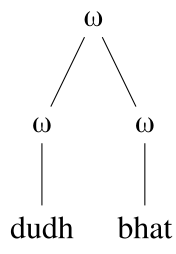




dúdh bhát

"milk" "rice"

When compounded, stress falls on the first syllable of the structure:



dúdhbhàt

"milk mixed with rice"

Bengali shares the same linguistic lineage as English, belonging to the Indo-European language family. As in English, Bengali compounds differ from noun phrases in both semantic and phonological structure: that is, the meaning of the compound, as well as the stress assignment, is distinct from that of the corresponding noun phrase. For example, *gɔŋga jɔl* (lit. "the Ganges" and "water" (generally)) differs from *gɔŋgajɔl*, (lit. "the water of the Ganges"). There are, however, differences in how Bengali compounds behave: individual compound constituents can be marked with case markers and/or inflection (e.g. *ghar-ebair=e*. "inside and outside the house").

Secondly, if the default prosodic structure of English compounds is indeed a recursive Phonological Word constituent, then more-fluent English speakers will, more often than not, endeavour to plan them as such. This brings us to our next hypothesis.

2. **Highly-fluent speakers will be able to access the correct prosodic frame for the compounds and phrases.**

While the claims made in the linguistic literature differ widely as to the structure of the bilingual mental lexicon, most seem to agree that initial specification of the target language (L2) occurs prior to phonological encoding. However, evidence that L2 phonological representations are present at the word-form encoding level comes from a number of chronometric studies (e.g., Colomé, 2001; Colomé & Miozzo, 2010) and ERP studies (e.g., Acheson et al., 2012; Rodriguez-Fornells et al., 2005). In Chapter 3, we discussed the advantages that WEAVER++ provides, in its comprehensive attention to the processes that prepare the phonological features of an utterance and in its ability to account for current experimental data. When applied to non-native speech production, WEAVER++ retains its original shape as a spreading activation model. Both L1 and L2 lemmas are activated in the non-native model, with nodes at each level that remind the processor to use the L2 language.

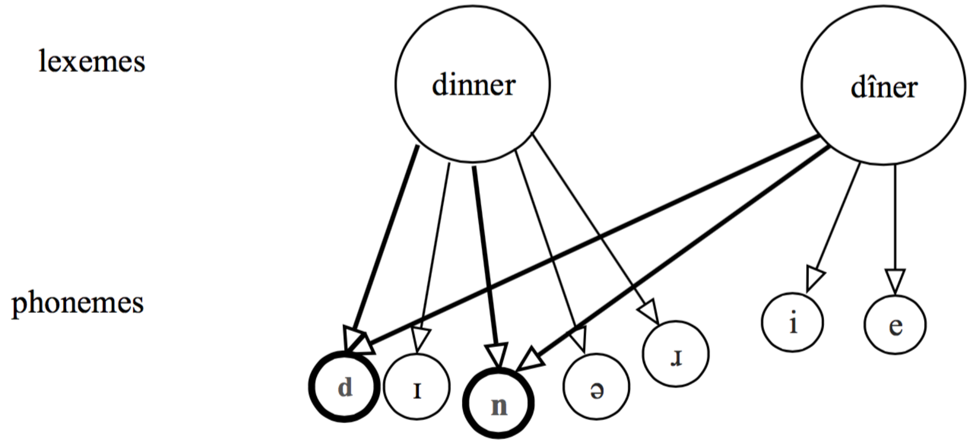


Figure 7.1: WEAVER++ for L2 Speakers

During phonological encoding, phonemes that are shared between the two languages receive extra activation; this explains why shared phonemes are planned faster than those that are not shared between languages.

Evidence for the L2 phonological encoding process, and the preparation of prosodic frames in particular, is sparse. Roelofs (2003) found that non-native English speakers behaved similarly to native speakers in the serial organisation of segments in phonological encoding. That is, they were able to plan segments from left to right, incrementally, without knowing the following segments of the English word. He also found a preparation effect with the initial segments were shared between the L1 and L2 language of the speaker: this, he countered, suggests that the phonological representations of similar segments are shared between the two languages. The data from these experiments not only supports WEAVER++'s organisation of phonological material in a native language, but shows that the process that prepares phonological material works the same in the L2 speakers.

Following this, we predict that if Bengali speakers are highly-proficient in English, they will have no trouble recognising the orthographic shape of compounds in the visual target and subsequently activating the necessary corresponding prosodic structures for compounds. Less-fluent speakers, however, will have multiple troubles: they may not recognise the orthographic target, and even if they do, they may not be able to plan it as a single prosodic unit. The discussion of fluency here has implications for the outcomes of this task. As discussed in Chapter 3 (§3.3.2), models of L2 speech often assume there is active interference by the L1 language during production. This assumption is tied to evidence from psycholinguistic tasks where bilinguals regularly exhibit the cost of speaking in the non-dominant language: for example, Roberts et al. (2002) found that non-native speakers of English uniformly named fewer pictures in the Boston Naming Test[2] than native monolingual speakers.

Furthermore, their naming latencies were significantly slower than those of native, monolingual English speakers. Gollan and Silverberg (2001) found that proficient Hebrew-English bilinguals made significantly more TOT retrieval failures than monolingual speakers. Crucially, the cost of speaking in one's L2 does not diminish when high levels of proficiency are reached (cf. a picture-naming task in Sholl et al., 1995, where the naming delay and asymmetry in concept retrieval persisted despite English-Spanish speakers being highly-fluent balanced bilinguals). Roelofs (2003) proposes that "a difference in fluency may mean that the serial ordering mechanism [shown above] works differently for the second than for the first language" (184). Support for this comes from Dell et al. (1997), who found that there was a marked difference in the types and number of errors between highly-practiced and less-practiced speech. This, Roelofs suggests, points to a reorganisation of the serial ordering mechanism; in multilingual speakers, this reflects in the practice that the speaker has had in the non-native language.[3]

Therefore, Experiment 1 will test two groups of Bengali speakers: those highly-proficient

[2]The Boston Naming Test is a frequently-used neuropsychological test of confrontation naming (where subjects are presented with an image they must name).

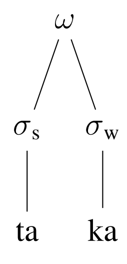
[3]The issues relating to practice, aquisition, exposure in the L2 are complex: while AoA[4] and AoE[5] are related, they are not necessarily balanced in the speaker. That is, speakers can be exposed to a wide range of vocabulary in the L2, but the AoA is not necessarily uniform across this exposure time. (Hanulová et al., 2011). Likewise, the amount of practice a speaker may have in their L2 is certainly variable.

in English and those with low proficiency. Errors will be especially important here: time outs (no response) or wrong items may suggest failures at the conceptual or lexical levels, both in recognition and production. Conversely, if the compound or phrase conditions elicit significant stress errors, this may signify trouble in the prosodic framing.

3. **Word stress will play a role in this experiment.**

In a 2011 paper, Chakraborty & Goffman tested whether proficiency influenced Bengali speakers' production of initial and final stressed words in English. They hypothesised that the iambic (finally-stressed) English words would present the most difficulty, and that less-proficient speakers would import the (incorrect) trochaic Bengali stress pattern into the frames of final stressed English words. Unsurprisingly, their speakers did exactly that: speakers with both high and low proficiency preferred to incorporate the trochaic Bengali pattern into their production of English words.

These results are wholly unremarkable: in fact, they were predicted by the linguist S.K. Chatterji nearly a century ago, in his *Origin and Development of the Bengali Language* (1926). Modern Bengali stress, he writes, "surrenders itself entirely" to sentence stress; as such, initial syllables predominantly carry heavy stress in both words and "sense groups" (281):



tákà

"unit of money"

In phonological phrases, this is almost always the left-most non-clitic word (or "important word" as Professor Chatterji calls it) (Hayes & Lahiri, 1991). And although stress can be phonetically-weak in Bengali, Hayes and Lahiri argue that stress must nevertheless be employed in the organisational framework of the language (much as in English) in order to explain specific phonological rules employed at the phrasal level. Further, while stress may be phonetically-indistinguishable in non-emphatic situations in Bengali, the employment of iambic stress patterns on trochaic Bengali words almost always results in a correction.[6] In conclusion, it is easy to see why iambic word structures would present the most difficulty for Bengali speakers in Chakraborty & Goffman's tasks.

Following this discussion of default word stress patterning in Bengali, we predict that Condition 4 of this task (disyllabic, finally-stressed morphosyntactic words) will elicit a large number of stress errors. Not only will speakers use trochaic stress, they may also exhibit longer reaction times in this condition. We argue that this will not be a reliable prediction of fluency; rather, the phenomenon will be present in both highly-fluent and less-fluent speakers alike. Reasons for this will be elaborated upon in the discussion in §7.2.5.

7.2 Experiment 5: Delayed Production of English Compounds in Highly-Fluent Bengali Speakers

The first experiment in this study consists of a delayed production task with a simple question-answer paradigm, identical to that in Experiment 1. Speakers in this group were highly-fluent non-native English speakers.

[6]For example, no native Bengali speaker would accept the English default pronunciation of *Àdítì* for *Ádìtì*.

7.2.1 Stimulus Selection

Stimuli were identical to those presented in Experiment 1. They are reproduced for easy reference in Table 7.1 below.

Table 7.1: Experimental Stimuli: Experiment 5

Condition 1: Compounds	Condition 2: Phrases	Condition 3: Initial Stress	Condition 4: Final Stress
daybreak	dark break	denim	default
daytime	dark time	debit	deceit
daylight	dark light	dagger	decree
groundhog	green hog	gravel	gazelle
groundwork	green work	griddle	gazette
groundnut	green nut	gravy	guitar
nightcap	nice cap	nibble	neglect
nightgown	nice gown	nickel	noblesse
nightclub	nice mare	nitrate	notate
lockjaw	long jaw	locker	lament
locksmith	long smith	locust	latrine
locknut	long nut	lodger	lapel
bathrobe	bad robe	banter	baboon
bathroom	bad room	ballad	balloon
bathtub	bad tub	basil	bamboo

7.2.2 5a: Delayed Production in Highly-Fluent L2 Speakers

7.2.3 Method

7.2.3.1 Design

The experimental design was the same as in Experiment 1. Participants were given written and spoken instructions in both English and Bengali.

7.2.3.2 Participants

Eighteen highly-fluent non-native English speakers between the ages of 18 and 22 took part in Experiment 5a. Participants were native speakers of Standard Colloquial

Bengali. They were recruited from Gokhale Memorial Girls' College in Kolkata, India, in March 2012. They were given course credit for their participation. They did not report any hearing, speaking, or reading impairments. All participants scored high on the proficiency tasks. No participant scored lower than Band 7 in the speaking task, and most were comfortably within Band 8 or 9 (see Appendix C for the band descriptors for this task). This experiment took place in a quiet room at Gokhale Memorial Girls College. Due to equipment constraints, the experimenter was seated at the far end of the same room, facing away from the participant.

7.2.4 Results

7.2.4.1 Latencies Analysis

Measurements were undertaken in the same way as in Experiment 1. Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Stress errors were not removed from analysis; these errors made up 48.8% of the error data. Other errors were also noticeably higher for this task, resulting in a loss of 10.4% of data. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error and removed from analysis. All data points (332 total) beyond two standard deviations from the mean were counted as outliers and removed. This, combined with the errors, resulted in a total loss of 19.9% of data.

The RT data were submitted to a linear mixed models analysis, in which naming latency (RT) was modelled as a function of the fixed effect factor, condition type (cond). Subjects and items were treated as random factors. The maximal model had condition as the fixed effect, random slopes for condition in the by-subject analysis, and random slopes for condition in a by-item analysis. This model[7] failed to converge. The next maximal model to converge contained a random slope for the by-subject analysis, and

[7] mla.maxstar <- lmer(rt ~ cond + (1+cond|sub)+ (cond|item)).

an intercept for the by-item analysis.[8]

This model converged; however, when submitted to residuals diagnostics, the model showed signs of a poor fit, particularly in the scatterplot analysis and histogram plot analysis (Figure 7.2). In an analysis of frequency effect (cf. §7.2.4), we found that total word frequency (CELEX) had an effect when treated for an interaction with condition. This indicated that the model needed to be updated to include the frequency interaction as a fixed effect.[9] This resulted in a better model fit; however, residuals analyses still showed signs of heteroscedasticity, particularly in the histogram and QQ-plot (cf. Figure 7.3). Therefore, a log-transform was run on the model: the Box-Cox transformation generated a log-likelihood of .383 (plot 1 in Figure 7.4). Furthermore, several outliers were established in leverage and fitted value plots. These outliers were removed, and the model was run with log-transform values of reaction times. This model [10]converged, and was also comfortably homoscedastic when fitted to residual plots, shown in figure 7.5.

Mean reaction times and percentage error rates for Experiment 5a are shown in Table 7.2. Mean latencies for the phrasal condition differed significantly (all ts > 3.03*) from all other conditions, while compounds and disyllabic words were statistically similar to one another. The final-stressed condition (Condition 4), however, did not reach significance in a pair-wise comparison to the phrasal condition. This will be discussed further in §7.2.5. Preparation time (beep) was significant when treated as a fixed effect in the analysis, but only generated significantly different naming latencies for the shortest preparation time in a pairwise comparison (all ts > -2.07) (Figure 7.6). There was also an effect (t= -3.35*) of experimental block on reaction times in this task: reaction times decreased significantly over the course of the experiment (all ts > 2.45*) (Figure 7.7). However, there was no interaction (t= 0.37) between condition and block.

[8] mla.maxplus3 <- lmer(rt ~ cond + (1+cond|sub)+ (1|item))

[9] mla.maxplus3 <- lmer(rt ~ cond + beep + cond*freq + (1+cond|sub)+ (1|item))

[10] mla.maxplus3 <- lmer(logrt ~ cond + beep +cond*freq+ (1+cond|sub)+ (1|item))

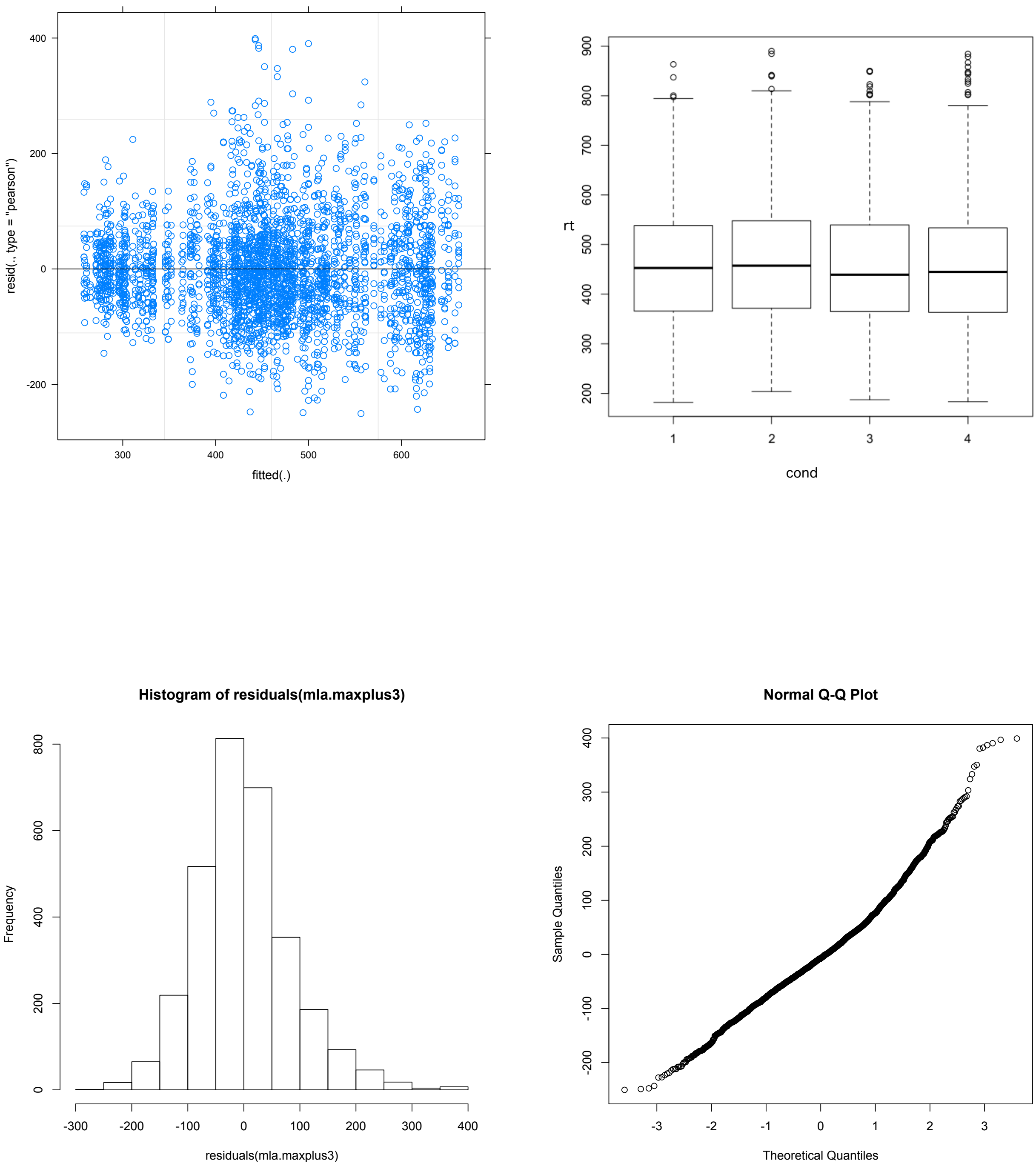


Figure 7.2: Residuals Analyses: Experiment 5a

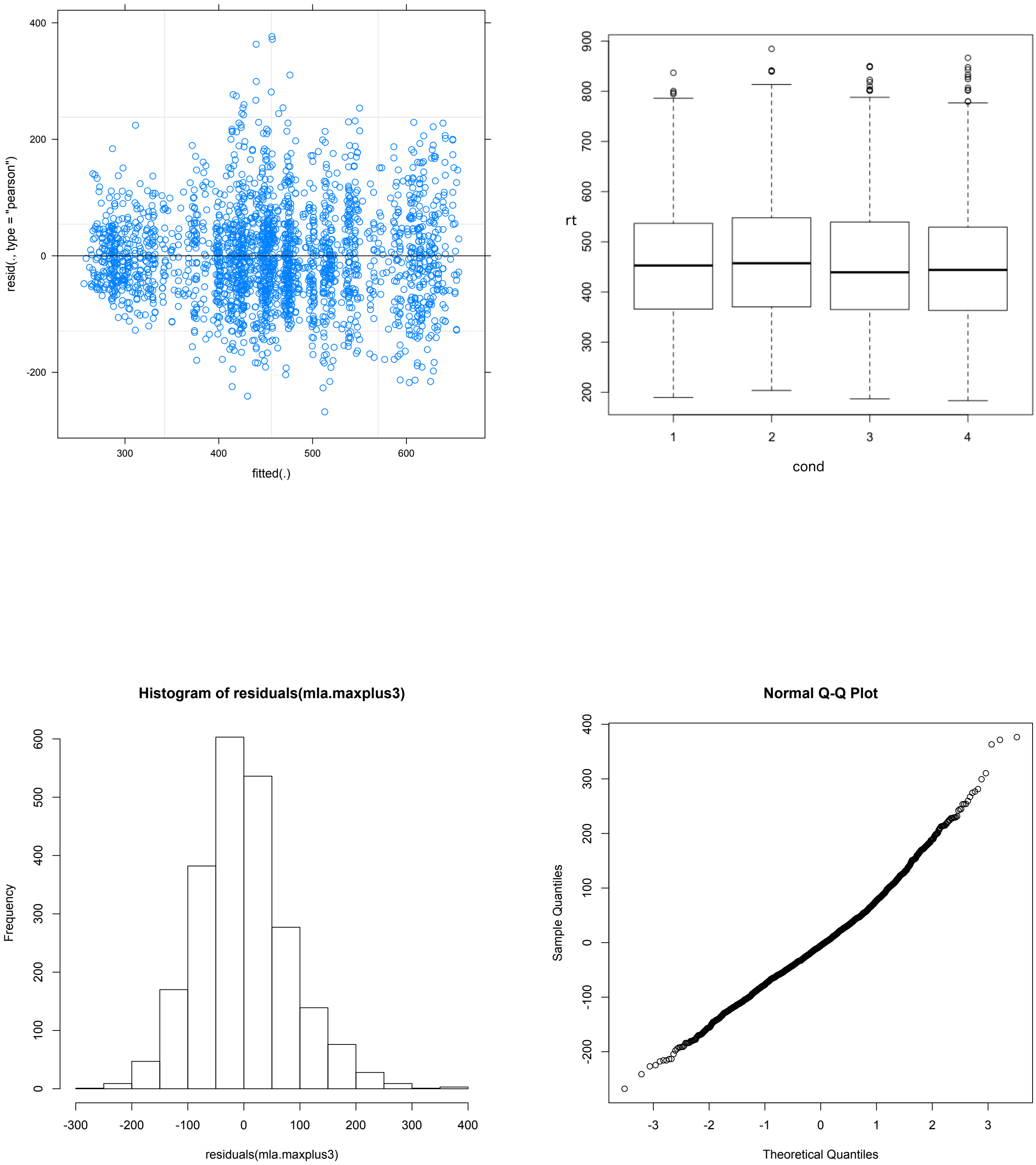


Figure 7.3: Residuals Analyses for Cond*Freq:Experiment 5a

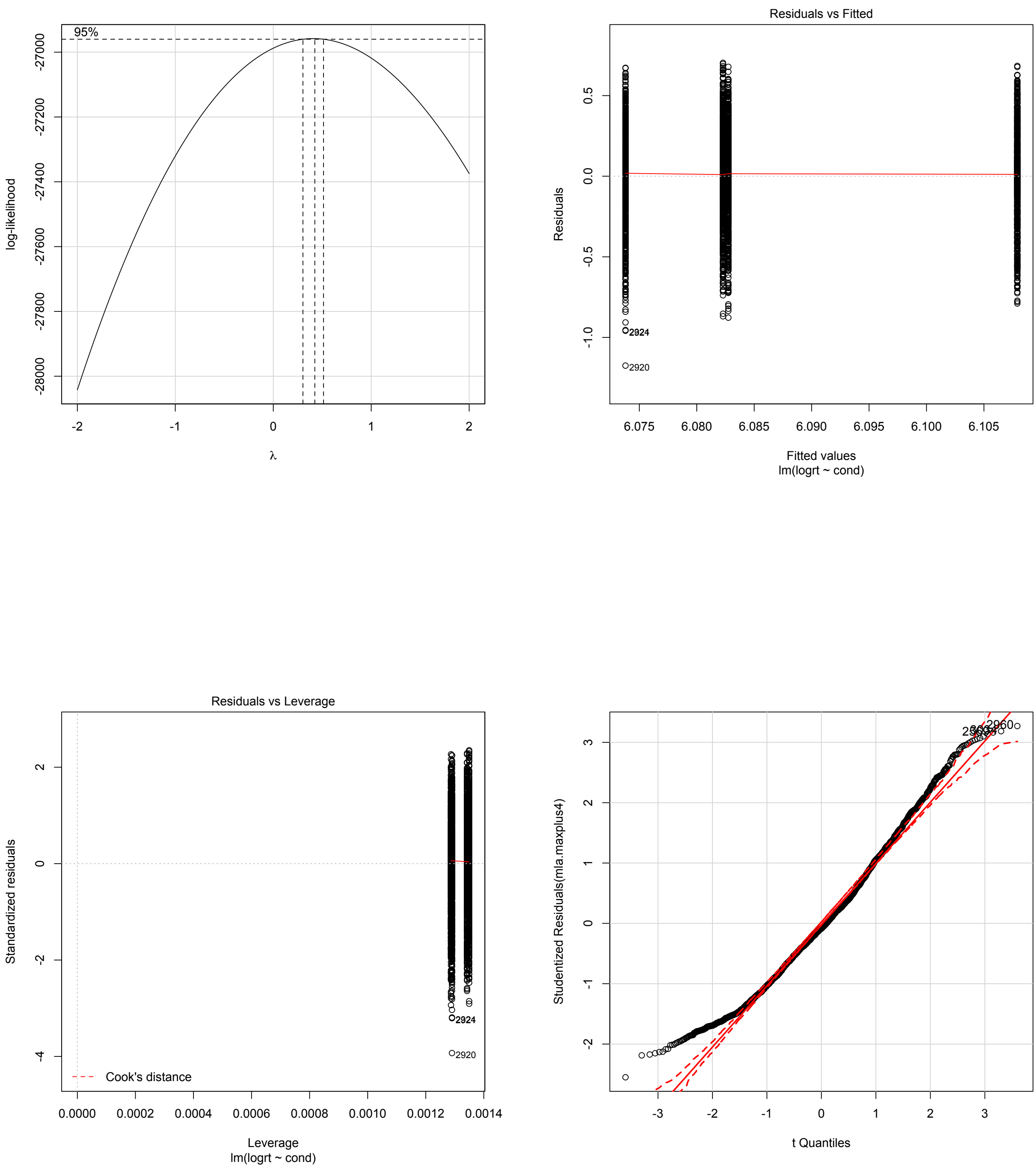


Figure 7.4: Residuals Transformations for Cond* Freq: Experiment 5a

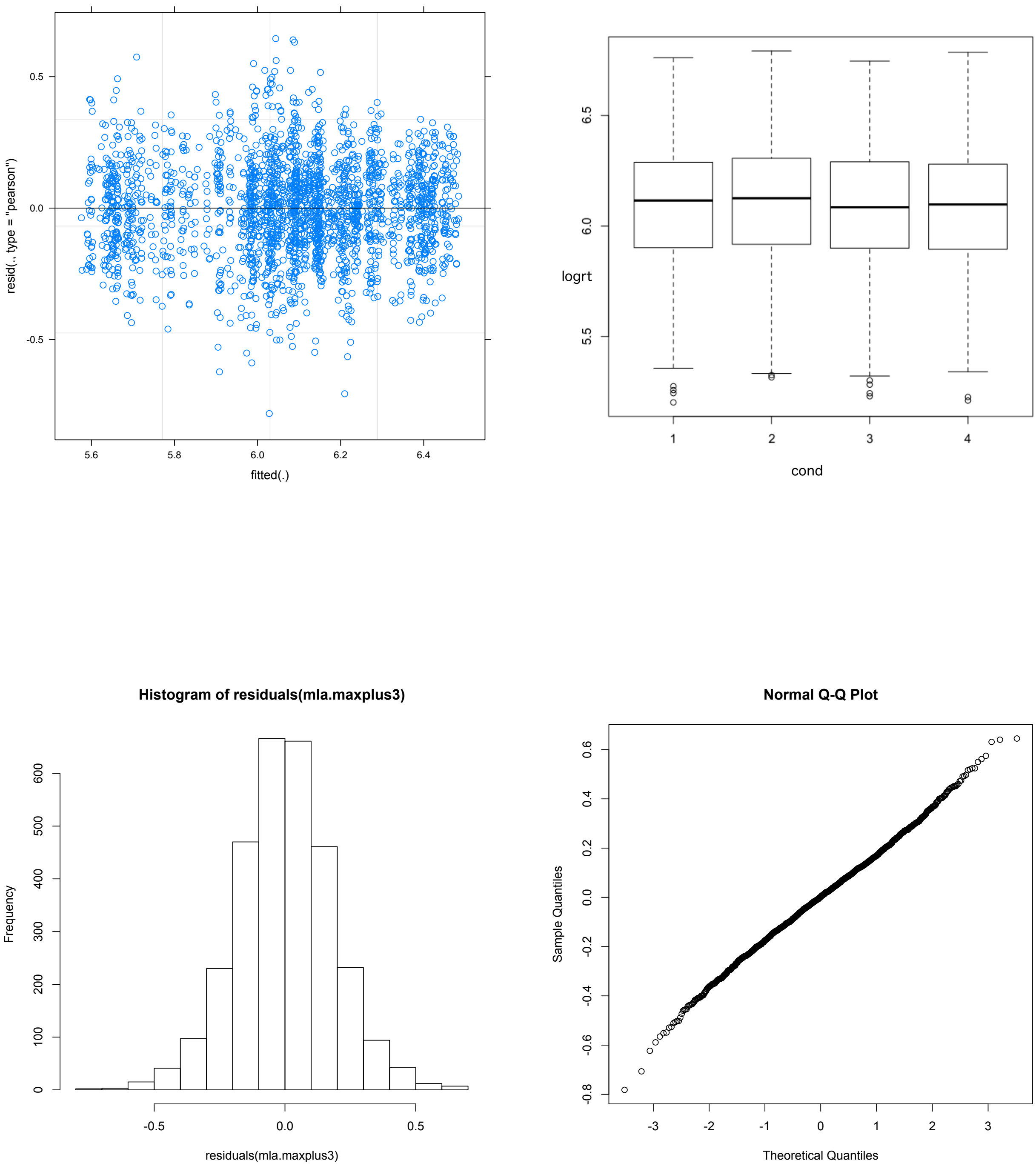


Figure 7.5: Residuals Log Analyses for Cond*Freq: Experiment 5a

Table 7.2: Linear Mixed-Effects Analyses of Transformed Data in Experiment 5a

rt	Estimate	SE	t-values
All Conditions:			
(Intercept)	6.116444	0.050041	122.23
Adj-N Phrases	0.030410	0.009969	3.05*
Initial Stress	-0.004320	0.011633	-0.37
Final Stress	-0.004389	0.011049	-0.40
1200 ms	-0.056522	0.008253	-6.85*
1400 ms	-0.058616	0.008265	-7.09*
Adj-N Phrase as Intercept:			
(Intercept)	6.146854	0.051057	120.39
Compounds	-0.030410	0.009969	-3.05*
Initial Stress	-0.034730	0.011474	-3.03*
Final Stress	-0.034799	0.010151	-3.43*
Disyllabic Initial as Intercept			
(Intercept)	6.11200	5.51102	110.91
Compounds	4.32003	1.16302	0.37
Adj-N Phrases	3.47302	1.14702	3.03*
Final Stress	-6.95705	1.059e-02	-0.01
Disyllabic Final as Intercept			
(Intercept)	6.119370	0.054113	113.09
Compounds	-0.003569	0.017488	-0.20
Adj-N Phrases	0.030129	0.017088	1.76
Initial Stress	-0.016121	0.017355	-0.93

Table 7.3: Mean Naming Latencies (in ms): Experiment 5a

	(1) Compounds	(2) Phrases	(3) Initial Stress	(4) Final Stress	Mean Lat. (ms)
PWds	1	2	1	1	
LexWs	2	2	1	1	
Syllables	2	2	2	2	
Beep Lat.:					
800 ms	475 (9.0%)	494 (13.6%)	464 (17.7%)	482 (38.6%)	479 (19.8%)
1200 ms	452 (10.3%)	463 (12.3%)	447 (16.3%)	446 (36.6%)	452 (18.9%)
1400 ms	439 (10.3%)	458 (12.7%)	450 (16.0%)	447 (35.3%)	448 (18.8%)
Mean Lat. (ms)	455 (10.1%)	471 (13.2%)	452 (17.0%)	450 (36.9%)	

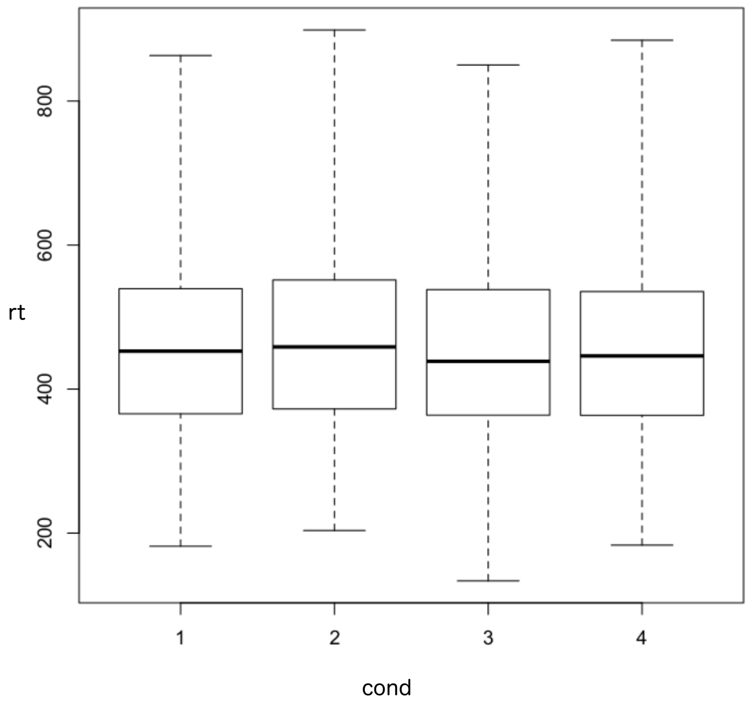


Figure 7.6: Mean RTs by Condition: Experiment 5a

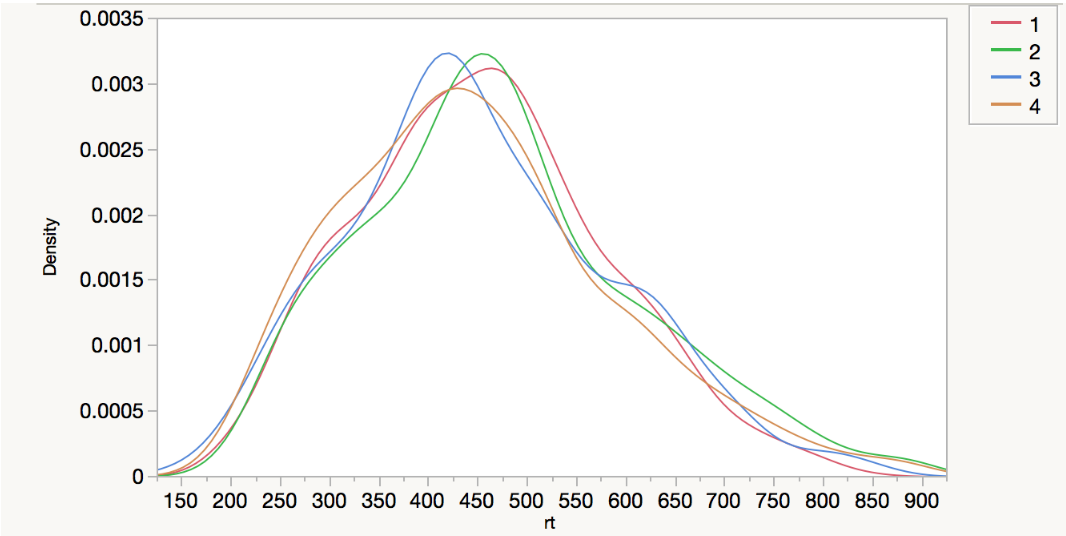


Figure 7.7: Densities of RTs by Condition: Experiment 5a

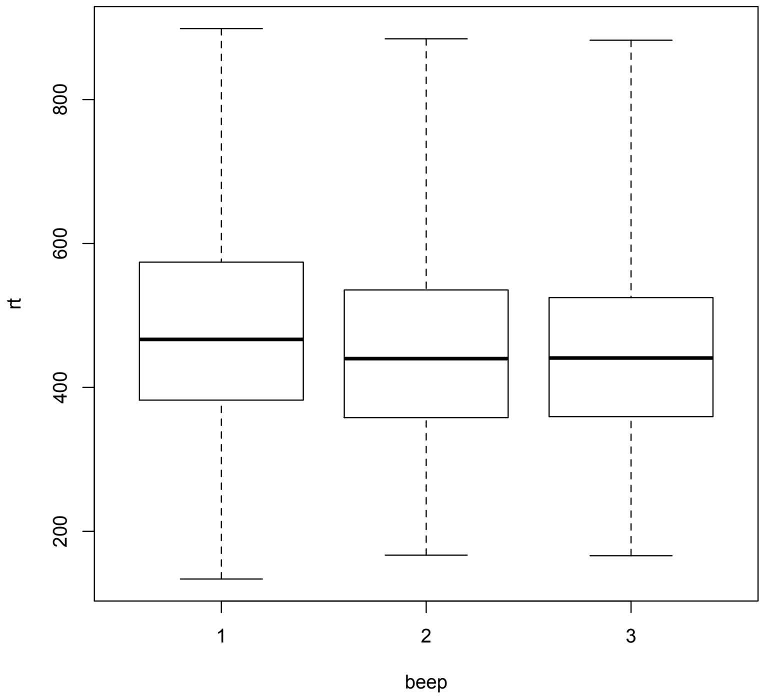


Figure 7.8: Mean RTs by Preparation Time: Experiment 5a

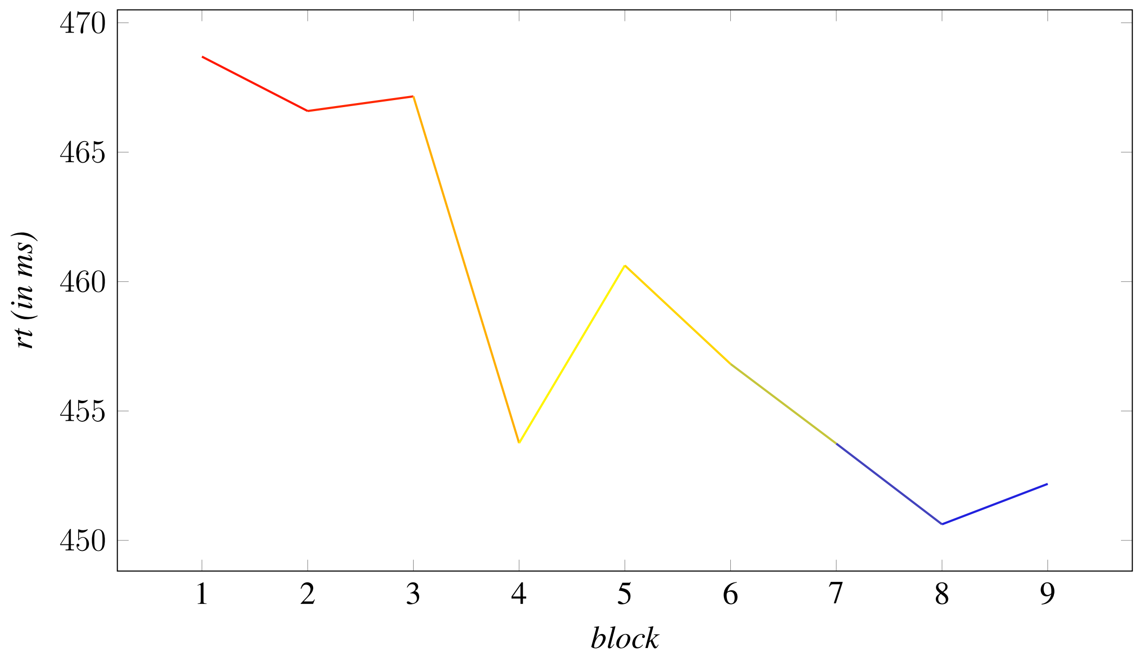


Figure 7.9: Mean RTs by Block: Experiment 5a

7.2.4.2 Errors

Errors were categorised as "time out" (the subject said nothing), "voice key error", "disfluency", "stress error", or "wrong item". The error rates were analysed proportionally using a generalised linear model (GLM) with a binomial distribution (link= logit). The analysis revealed a main effect of condition on error rates (χ^2= 66.56, p= <.0001*). Overall, these speakers made significantly more errors than the native speakers in Experiment 1. This was not wholly unpredicted, considering the increase in difficulty related with L2 planning. Condition 1 resulted in the fewest errors, while Condition 4 elicited over 300 errors (out of 900 total trials) (cf. Figure 7.10 below) Aside from stress errors, "wrong items" were the most common error for speakers to make in this task, followed by disfluencies and time outs. Errors were higher for the shortest preparation time (800 ms), but preparation time did not have a significant effect on the error rates (χ^2= 3.31, p= 0.19). There was a weak effect of block on error rate (χ^2= 82.43, p= <.046*): with blocks 1 and 2 resulted in slightly more errors (χ^2= 6.00, p= .01 and χ^2= 4.13, p= 0.041* respectively).

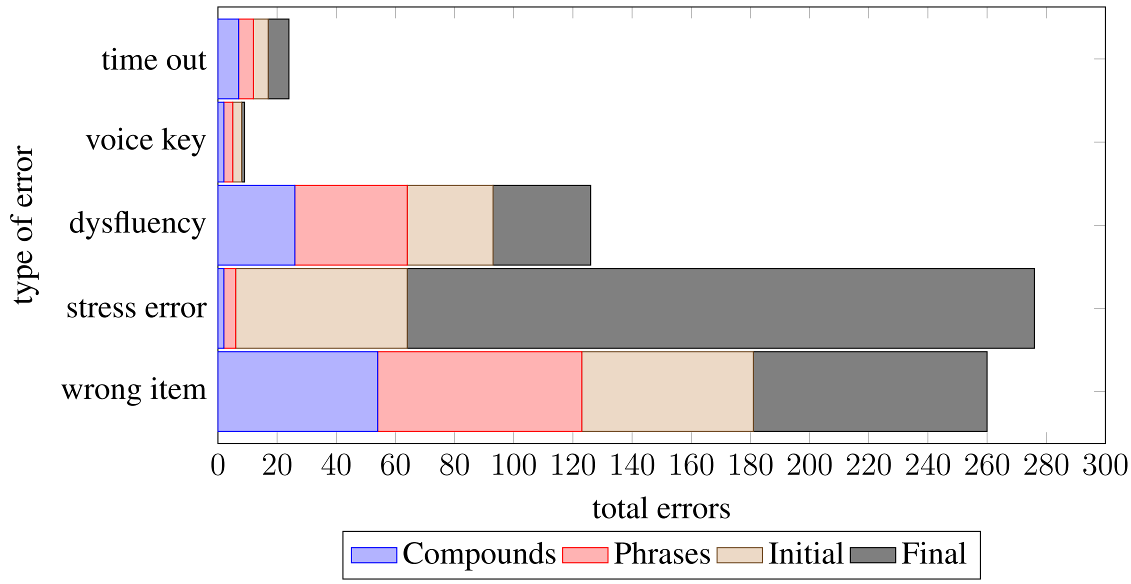


Figure 7.10: Total Errors: Experiment 5a

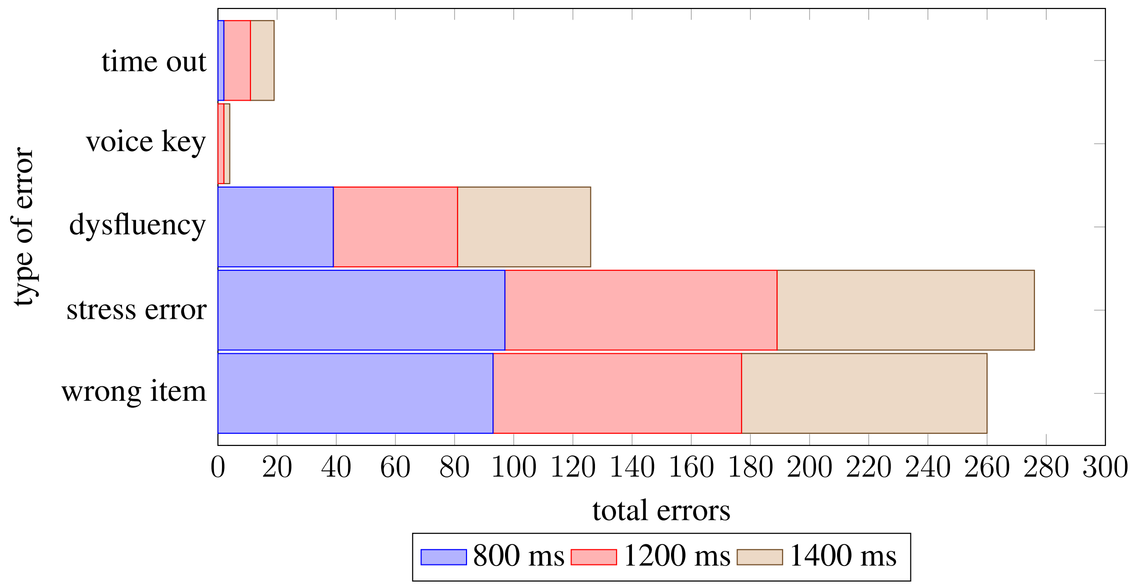


Figure 7.11: Errors by Preperation Time: Experiment 5a

7.2.4.3 Frequency Analyses

The RT data were again submitted to a linear mixed models analysis, in which frequency and condition were treated as the fixed effect factors. Once again, the maximally-appropriate structure that converged presented random intercepts and slopes for subject factors and a random intercept only for the between items factor.[11] The models converged.

Total (CELEX) word frequency had an interaction on condition for reaction times in this experiment: (Estimate= -0.0014097, SE= 0.0006575, t= -2.14*) and consequently, on compound latency (Estimate= -0.6486, SE= 0.3167, t= -2.048*). Neither the frequency of the first (Estimate= -0.003793, SE= 0.010564, t=-0.359) nor second (Estimate=-0.001681, SE=0.006024, t= -0.279) morpheme of the compound elicited a significant effect on rt. Likewise with the adjective (Estimate= 1.49204, SE= 1.35202, t= 0.011) and noun (Estimate= 3.62004, SE= 7.30103, t= 0.05) morphemes of the phrasal condition.

An analysis for an interaction effect between condition and word length did not reach significance (Estimate= 0.04816, SE= 0.24747, t= 0.195) but overall, reaction times were significantly longer for longer words (Estimate= 0.005929, SE= 0.002363, t= 2.51*): latencies for words containing 10 letters were around 10 ms longer than those for words containing only 5 letters (see Figure 7.11 below). This will be discussed further in §7.2.5.

[11] e.g.: rt ~ cond + freq + (1 + cond | sub) + (1 | item).

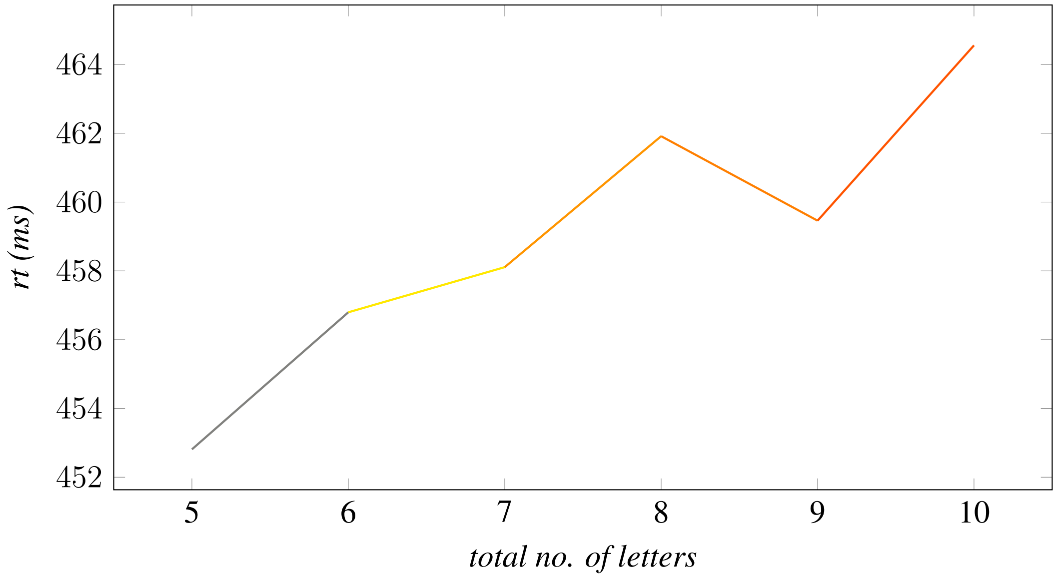


Figure 7.12: Mean RTs by Word Length: Experiment 5a

7.2.5 5b: Delayed Production in Less-fluent L2 Speakers

In §7.1.1, we discussed possible effects of fluency would have on this task. In order to address these issues, this experiment was also run on a second group of participants: eighteen native Bengali speakers with low fluency in English.

7.2.5.1 Participants

Eighteen native speakers of Standard Colloquial Bengali were recruited for this task. Unlike those speakers in Experiment 5a, these subjects were in the Bengali language track at Gokhale Memorial Girls College and therefore, had low proficiency in English. They consistently scored low on both the reading and speaking proficiency tasks, averaging just 4.5 (out of 9 possible Bands) on the speaking task (Band descriptors can be found in Appendix C). They were recruited from the College in Kolkata, India, in March 2011, and given course credit for their participation. They did not report any hearing, speaking, or reading impairments. This experiment took place in a quiet room at Gokhale Memorial Girls College. Due to equipment constraints, the experimenter was seated at the far end of the same room, facing away from the participant. Prior to the experiment, subjects were given written and verbal instructions in Bengali, and were assisted during the practice session by a native Bengali speaker.

As predicted, all of the latencies were longer for the less-fluent speakers: they took, on average, 70-80 milliseconds longer to initiate speech. The effect of preparation time had a significant effect on naming latencies with the shortest preparation time (t= 4.133*). However the difference in production latency between phrases and compounds was not significant with this group. Results for this group also revealed a high number of errors: including stress errors, errors for this group added up to nearly 38% of the data. Condition 4 (final stress) resulted in a very large number of stress errors, as expected from the Bengali stress rule. Incorrect items were the most common errors for this group, with around 600 total incorrect responses given to the stimuli. 12 (out of the 18) participants had a greater number of incorrect items than any other error.

Table 7.4: Naming Latencies (in ms): Experiment 5b

	(1) Compounds	**(2) Phrases**	**(3) Initial Stress**	**(4) Final Stress**	**Mean Lat. (ms)**
PWds	1	2	1	1	
LexWs	2	2	1	1	
Syllables	2	2	2	2	
Beep Lat.:					
800 ms	622 (19.5%)	604 (18.1%)	586 (16.3%)	560 (49.3%)	594 (26.5%)
1200 ms	543 (18.2%)	591 (17.3%)	553 (16.2%)	535 (48.4%)	556 (17.6%)
1400 ms	521 (18%)	553 (18.6%)	583 (16.4%)	562 (45.3%)	554 (19.2%)
Mean Lat. (ms)	562 (18.0%)	582 (16.3%)	573 (19.3%)	552 (47.2%)	

Table 7.5: Linear Mixed-Effects Analyses of RT*Preparation Time Data: Experiment 5b

Term	Estimate	SE	t value
(Intercept)	581.474	56.197	10.347
Adj-N Phrases	17.213	17.995	0.957
Disyllabic Initial	8.554	12.282	0.697
Disyllabic Final	-6.460	12.410	-0.521
1200ms	-33.898	9.761	-3.473*
1400ms	-36.262	9.783	-3.707*

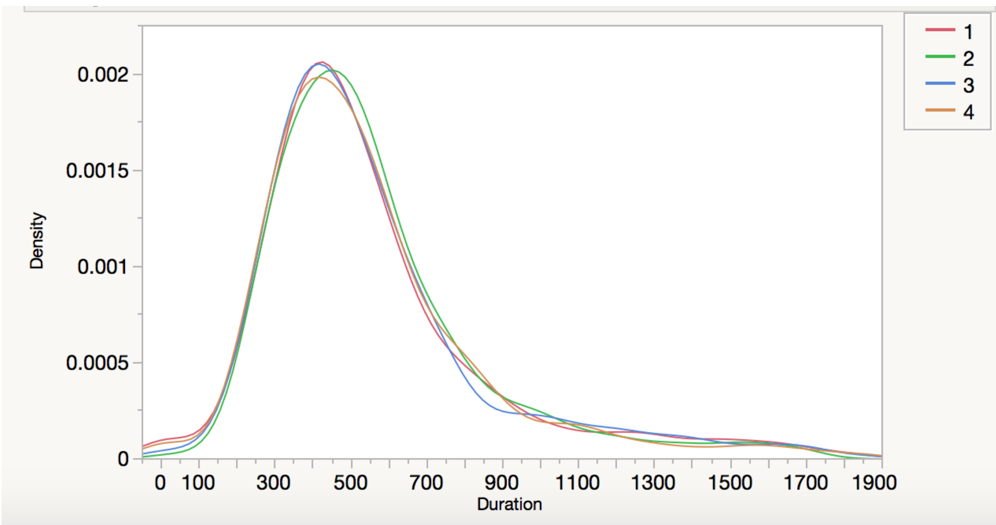


Figure 7.13: Densities of RT by Condition: Experiment 5b

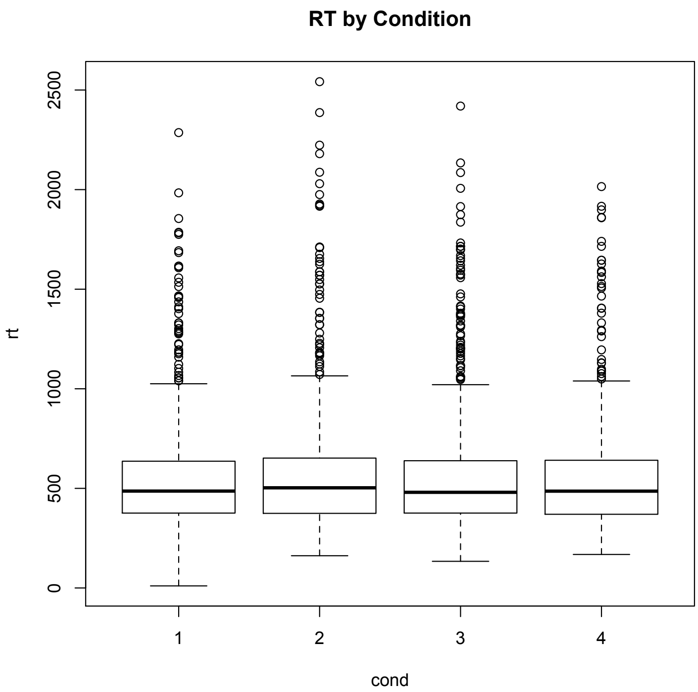


Figure 7.14: Mean RTs by Preparation Time: Experiment 5b

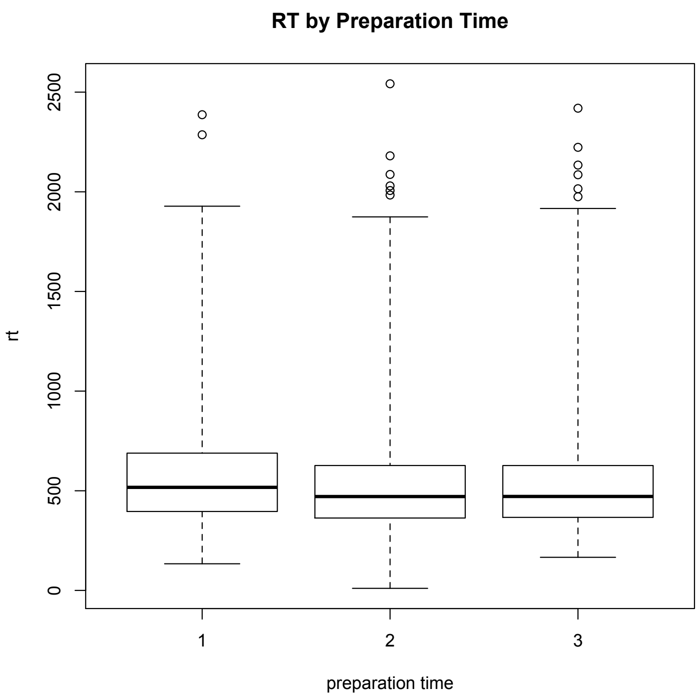


Figure 7.15: Mean RTs by Preparation Time: Experiment 5b

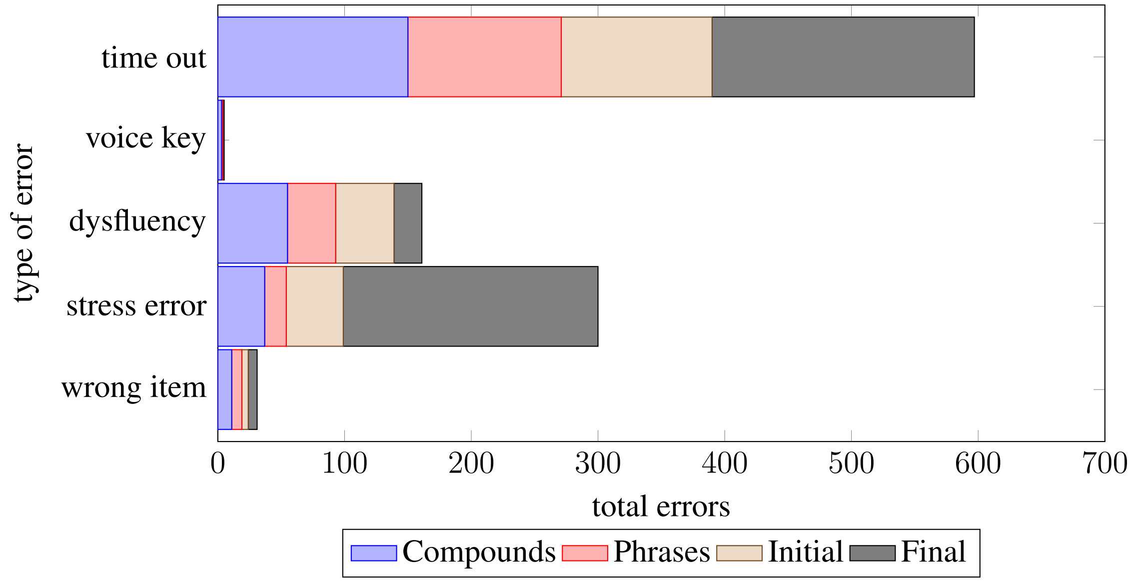


Figure 7.16: Total Errors: Experiment 5b

7.2.6 Discussion

These experiments yielded some interesting results. The highly-fluent speakers (Experiment 5a) behaved similarly to the native speakers in regards to the planning of the compound and phrasal conditions. Compounds elicited naming latencies that were an average of 15 ms shorter than adjective-noun phrases (t= 3.05*). While overall latencies were not significantly longer than those of the native speakers (compare 455 ms for native speakers to 457 ms for non-native speakers), the size of the effect was much smaller than that seen for the native speakers in Experiment 1 (which yielded a 40 ms. difference). This is in line with studies that find evidence of inhibition in L2 naming tasks (Dimitropoulou et al., 2011; Nakayama et al., 2012; Ando et al., 2014).

Compounds elicited similar latencies to those of initial and final stress (all ts < 0.37); however, finally-stressed disyllabic simple words (Condition 4) did not significantly differ from any of the other conditions in a pairwise analysis (all ts < 1.7). This condition also resulted in a significant number of errors: over 300 errors out of 900 total trials. As predicted in the hypotheses for this experiment, both of these results are easily explained by the existence of the Bengali stress rule. If we return to WEAVER++, we look at how this might happen at the phonological encoding stage.

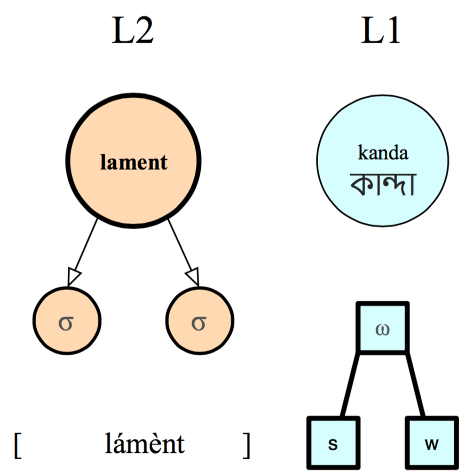


Figure 7.17: Incorrect Stress Errors in the L2 Encoding Process

We know from picture-word interference studies (cf. Costa et al, 1999) that the L1 lemmas are often activated alongside L2 lemmas in non-native speakers (in Figure 7.17 above, the Bengali lemma *kanda* ("to weep, lament") is activated alongside the target *lament*). We also know that shared phonemes speed up naming in these speakers, suggesting that speakers draw from a shared phonological pool. The Phonological Word frames specify the syllables and stress patterns for the arrangement of these phonemes. Roelofs & Verhoef (2006) assume that certain aspects of L2 phonological encoding cannot be shared (as the inventory of phonemes is): languages differ in terms of rules of syllabification and stress assignment.

While English words can differ in stress placement, Bengali is overwhelmingly stress-initial. The speaker may be less familiar with the word (*lament*), so while they are able to recognise the orthographic representation and activate the associated phonology, word familiarity does not extend to stress placement: and indeed, we saw evidence for the effect of total word frequency in this task (t= -2.14*). Recall that WEAVER++ also places a "reminder node" to use the L2 in the working memory of the non-native speaker: the lack of familiarity with the word *lament* coupled with the effort of speaking in the

L2 language results in the default Bengali prosodic frame being assigned to the word: *lámènt*. Chakraborty and Goffman (2011) assign this phenomenon to "nonnative speakers appear[ing] to overgeneralise the Bengali trochaic stress pattern, which in turn suggests an L1-to-L2 transfer of lexical stress..." (211). Such a statement suggests a conscious effort to apply trochaic stress to all English words. There is no transfer of lexical stress, nor is there an overgeneralisation of stress patterns: the process is simply defaulting to the stored prosodic shape for disyllabic words in the Bengali speakers.

The error analysis also revealed some interesting differences between the native and non-native speakers: aside from stress errors, the most frequent error that the non-native speakers made was "wrong item". It is noteworthy that these errors often shared initial phonological segments, as those reproduced in example 7.1 below:

(7.1) a. *bell* for *ballad*

b. *lemon* for *lament*

c. *bad road* for *bad robe*

This lends support to the incremental activation and planning of segments in both native and non-native speakers of English. The increase in task difficulty was evident not only from the number of errors, but also from other measurements: we found an overall effect of word length (t= 2.51*) on reaction times, as well as an interaction between condition and simple word frequency. This was the first experiment in this study to elicit a frequency interaction effect of any type, and therefore deserves further investigation. Duyck et al. (2008) also elicited a frequency effect in a visual word recognition task using Dutch-English bilinguals. They observed a frequency effect in L2 reading that was nearly twice as large as that in L1 reading: this, they argue, relates lexical access to learning process (852). Shi (2015) also found evidence of a frequency effect in L2 speakers; furthermore, this effect increased as the fluency of the speaker decreased. Shi also reported that speakers in their task tended to replace the target words with higher frequency words, e.g. *lure* for *lore*. If we return to the errors the Bengali speakers made

(Example 7.1), we see that we have generated evidence that agrees with this finding:

Table 7.6: Frequencies and Errors: Experiments 5a and 5b

Correct Item	Frequency	Error Item	Frequency
ballad	50	bell	745
lament	23	lemon	278
robe	257	road	4458

These findings fit with the predictions of the WEAVER++ model of L2 production: if the L2 word has less use (or has been learned later in life), it will likely receive less activation; if the L1 and L2 languages are unbalanced to begin with, all L2 lemmas are already receiving less activation. The activation effect is therefore magnified in non-native speakers, resulting in words with lower frequency generated significantly longer reaction times in our data.

In Experiment 5b, we found very different results for the delayed naming task. The less-fluent native speakers were not able to distinguish between any of the four conditions. Although the reaction times superficially showed a difference between compounds and phrases in a comparison of ms, condition type did not exhibit a significant effect on naming latency in the LME model analysis (all ts < .9). Errors for these speakers were also very high, with 33% of all trials resulting in some form of incorrect answer. Stress errors were very high, but the second most common form of error for these speakers was a time out (null response). Returning to the predictions made in §7.1.1, we have evidence here that relative language fluency does indeed have an influence on the results. These findings replicate those of other language production tasks, e.g. those found in Christoffels et al., (2006); Hoshino & Kroll, (2008); Poarch & Van Hell, (2012). Not only did less-fluent speakers generate a large number of errors and significantly-longer reaction times, they also were unable to differentiate between the compound and phrasal conditions. Now that we have established the importance of fluency for our task, we can continue with the next experiment: testing the planning of compound words and adjective-noun phrases in online conditions.

7.3 Experiment 6

7.3.1 Introduction

Experiment 5 provided evidence that advanced non-native speakers are able to recognise compounds as single prosodic units when they are given time to plan their utterances. However, we do not know the effect that online planning will have on this process. Our hypotheses will concentrate on the demands associated with second language production and the removal of planning time.

7.3.2 Hypotheses

1. **The demands of the online production task will be noticeable for even these highly-fluent non-native speakers.**

Very little evidence exists for production in online task conditions in L2 speakers. Mehnert (1998) obtained evidence that L2 speakers who were given time to prepare their responses to a series of tasks produced "significantly more fluent and more accurate speech with higher lexical density" than those who were required to answer immediately (99). Ahmadian et al. (2012) found that, when performing a task with an unstructured storyline (that is, a task that has a beginning, middle, and end) under online planning conditions, L2 speakers scored lowest in all three of their test areas: fluency, accuracy, and complexity (41).

We discussed the relationship between working memory and speech production in the results of Experiment 2 (§5.3.7): as the production processes are preparing their corresponding representations, they are simultaneously being maintained by limited-capacity buffers (cf. Martin & Freedman, 2001). During phonological encoding, the unfolding phonological segments are compiled and stored in a buffer that contains the phonological output forms. The demands of online planning are seen in the limitation of planning scope and the increase in error rates. Speaking in a non-native language also increases the demands on working memory: if representations in both languages are being activated at different stages, then it follows that these representations are also being held in the working memory buffers (cf. Baddeley, 1986, 2003). These memory buffers, which are already taxed by the limited preparation time of the online task conditions, may be further limited by the effort of maintaining information in both the L1 and L2. As well as an increase of errors, we must entertain the possibility that these speakers may not be able to prepare the correct L2 prosodic frames for our experimental conditions in this task.

2. **If speakers are able to plan correct prosodic frames in this task, then compounds will retain their recursive word shape.**

We saw in Experiment 2, that native English speakers continued to plan compounds as single prosodic units even when planning time was removed. This resulted in significantly different naming latencies for the compound and phrasal condition, as speakers only planned the adjective of the phrase in advance of utterance initiation. Additionally, speakers waited until the entire compound word was phonologically-prepared before initiating speech, suggesting that speakers prefer to plan well-formed Phonological Words even at the cost of utterance initiation speed.

If the L2 speakers are able to encode the correct English frames for the compounds and phrases, then we predict that the compounds will retain the shape we have seen in the past five experiments: that of a recursive Phonological Word. Accordingly, the naming latencies for the adjective-noun phrases will only reflect the preparation of the first prosodic unit (the adjective). Therefore we predict that naming latencies for the compounds will be similar to those of the disyllabic simple words, and the latencies for the phrases to those of the monosyllabic simple words, as in Experiment 2 (§5.3.1)

7.3.3 Stimulus Selection

Stimuli were identical to those presented in Experiment 2. They are reproduced in Table 7.7 below.

Table 7.7: Stimuli: Experiment 6

Condition 1: Compounds	**Condition 2: Phrases**	**Condition 3: Disyllabic**	**Condition 4: Monosyllabic**
daybreak	dark break	denim	dame
daytime	dark time	debit	date
daylight	dark light	dagger	daze
groundhog	green hog	gravel	grease
groundwork	green work	griddle	greed
groundnut	green nut	gravy	greet
nightcap	nice cap	nibble	nine
nightgown	nice gown	nickel	Nile
nightclub	nice mare	nitrate	ninth
lockjaw	long jaw	locker	lob
locksmith	long smith	locust	log
locknut	long nut	lodger	loft
bathrobe	bad robe	banter	bag
bathroom	bad room	ballad	band
bathtub	bad tub	basil	bat

7.3.3.1 Design

The experimental design was the same as in Experiment 1. Participants were given written and spoken instructions in both English and Bengali.

7.3.3.2 Participants

Twenty-one highly-fluent non-native English speakers between the ages of 18 and 22 took part in Experiment 6. Participants were native speakers of Standard Colloquial Bengali. They were recruited from Gokhale Memorial Girls' College in Kolkata, India, in March 2012. They were given course credit for their participation. They did not report any hearing, speaking, or reading impairments. All participants scored high on the proficiency tasks. No participant scored lower than Band 7 in the speaking task, and most were comfortably within Band 8 or 9 (see Appendix C for the band descriptors for this task). This experiment took place in a quiet room at Gokhale Memorial Girls College. Due to equipment constraints, the experimenter was seated at the far end of the same room, facing away from the participant.

7.3.4 Results

7.3.4.1 Latencies Analysis

Measurements were undertaken in the same way as in Experiment 2. Errors resulted in a loss of 9.5% of data. Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error. All data points beyond two standard deviations from the mean were counted as outliers and removed. The RT data were submitted to a linear mixed models analysis, in which naming latency (RT) was modelled as a function of the fixed effect factor, condition type (cond). Subjects and items were treated as random factors. The maximal model had condition as the fixed effect, random slopes for condition in the by-subject analysis, and random slopes for condition in a by-item analysis. This model[12] failed to converge. The next maximal model to converge contained a random slope for the by-subject analysis, and an intercept for the by-item analysis.[13]

This model converged; however, when submitted to the residuals diagnostics, the data showed evidence of heteroscedacity (see Figure 7.8). Therefore, a log transform was conducted. Based on the Box-Cox calculation (Plot 1 in Figure 7.19), the data was transformed to a power of -0.06. This resulted in a much more homoscedastic distribution in the scatterplot, save for two outliers. These outliers were identified and removed. The result of the transformation and outlier-removal was a very homoscedastic distribution of residuals (Figure 7.20). The plot points are evenly distributed in the scatterplot, and the QQ-plot in 7.20 is much straighter. The new model converged and its values are presented in Table 7.6.

Mean reaction times and percentage error rates for Experiment 6 are shown in

[12] mla.maxstar <- lmer(rt ~ cond + (1+cond|sub)+ (cond|item)).
[13] mla.maxplus3 <- lmer(rt ~ cond + (1+cond|sub)+ (1|item))

Table 7.8. The mean latencies for compounds and disyllabic words were statistically similar to one another, as were the latencies for adjective-noun phrases and monosyllabic words. Furthermore, compounds and disyllabic words differed significantly from phrases and monosyllabic words. Reaction times for Block 1 were significantly (Estimate= -21.5558, SE= 10.6148, t= 2.031*) slower than Block 2, 3 and 4, and rose again in Block 5 (Figure 7.21).

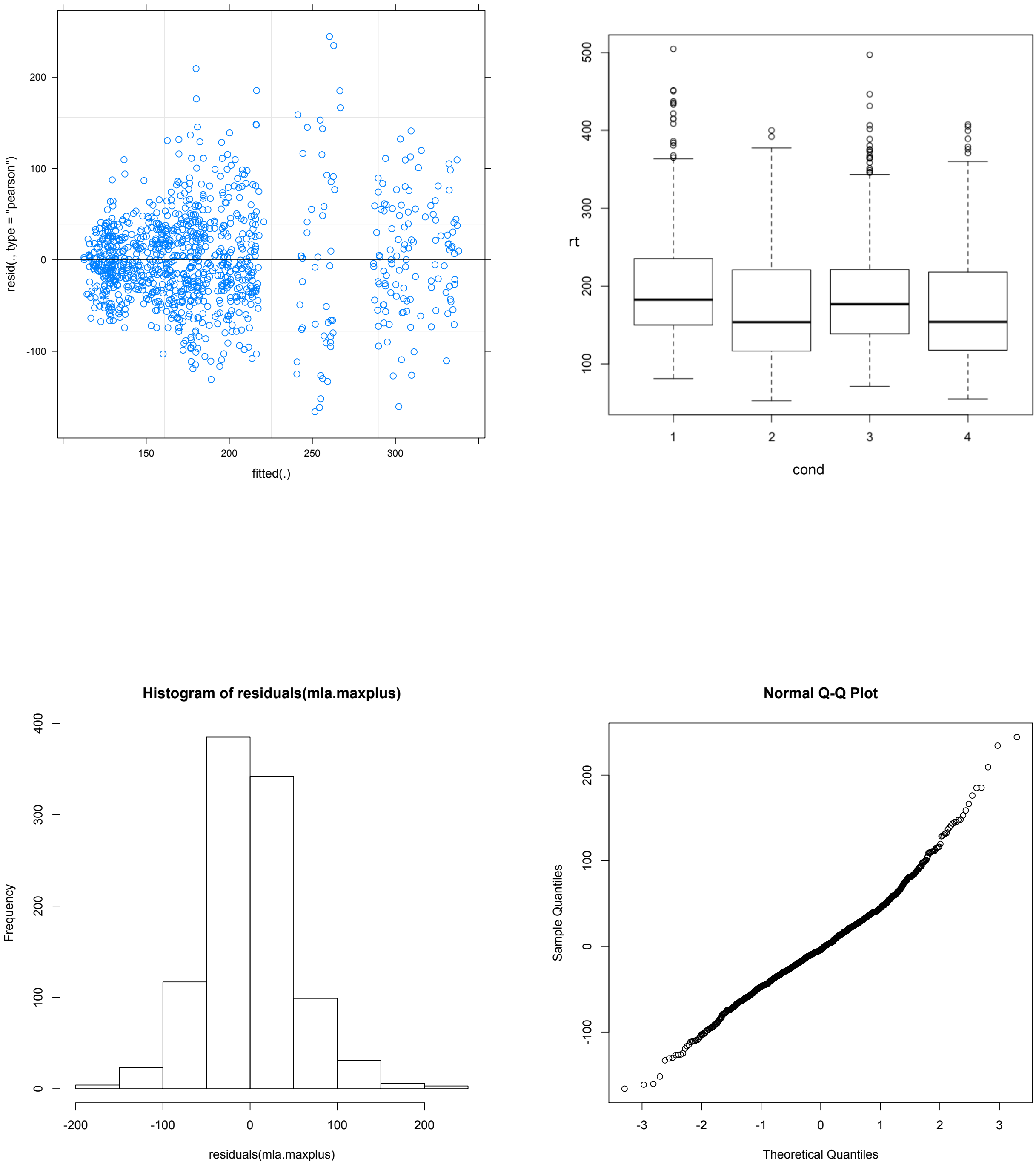


Figure 7.18: Residuals Analyses: Experiment 6

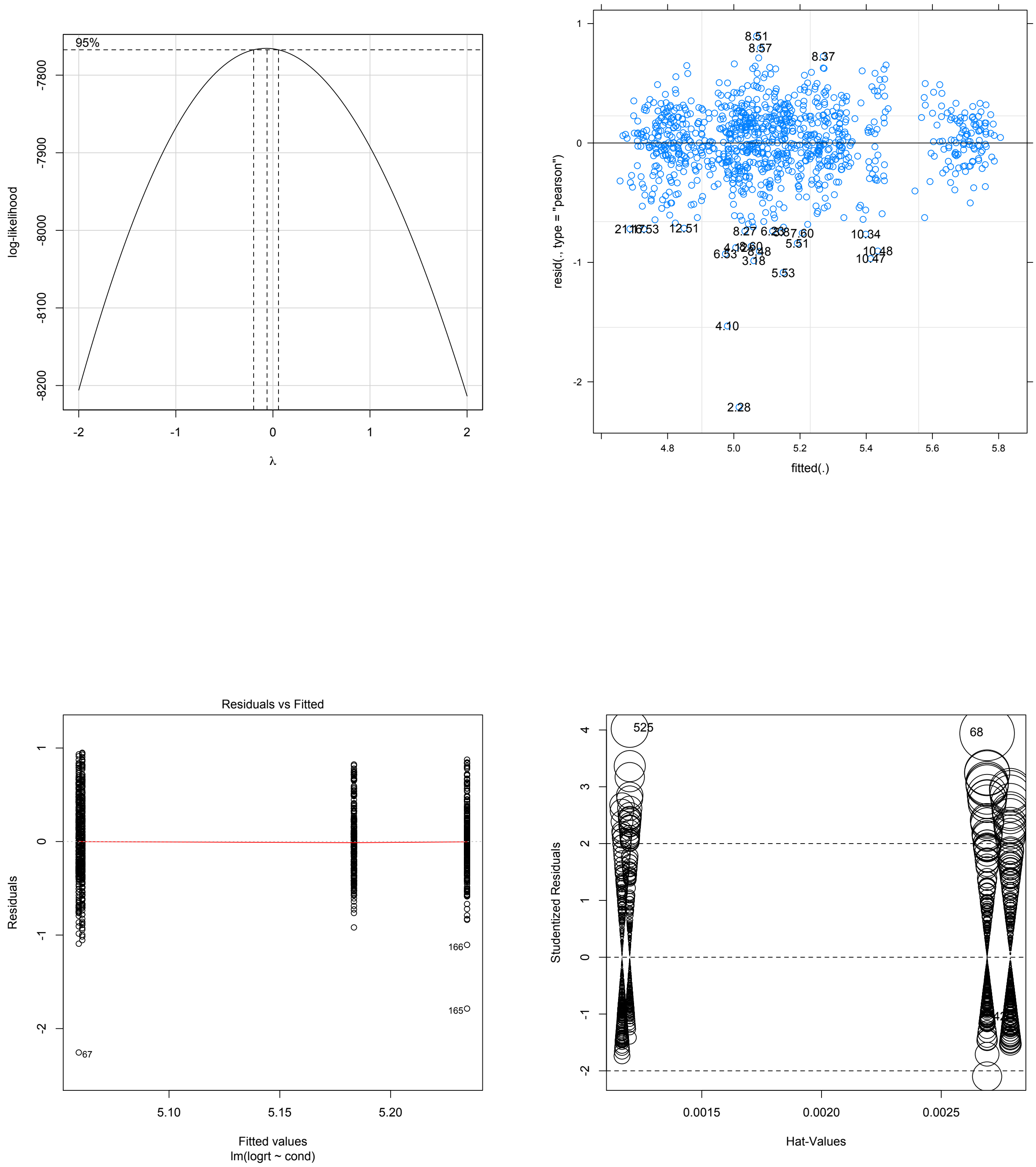


Figure 7.19: Residuals Log Transformation: Experiment 6

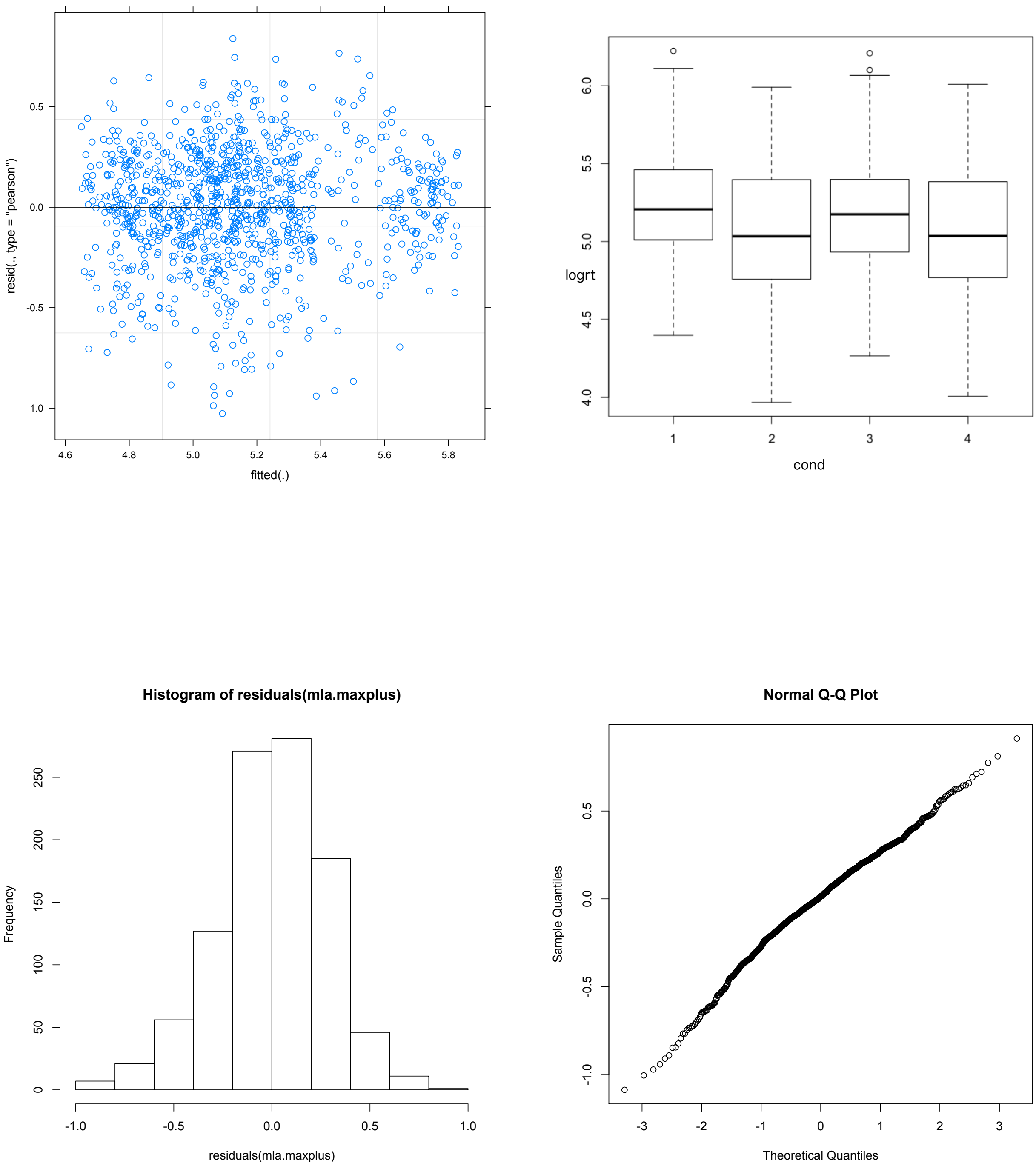


Figure 7.20: Transformed Residuals Analyses: Experiment 6

Table 7.8: Linear Mixed-Effects Analyses of Transformed Data: Experiment 6

	Estimate	**SE**	**t values**
All Conditions:			
(Intercept)	5.25672	0.05739	91.60
Adj-N Phrases	-0.19602	0.04645	-4.22*
Disyllabic Initial	-0.06853	0.03872	-1.77
Monosyllabic	-0.19243	0.03908	-4.92*
Adj-N Phrase as Intercept:			
(Intercept)	5.060695	0.072747	69.57
Compounds	0.196025	0.046446	4.22*
Disyllabic initial	0.127499	0.039097	3.26*
Monosyllabic	0.003594	0.032909	0.11
Disyllabic as Intercept:			
(Intercept)	5.18819	0.06338	81.86
Compounds	0.06853	0.03872	1.77
Adj-N Phrases	-0.12750	0.03910	-3.26*
Monosyllabic	-0.12390	0.03550	-3.49*
Monosyllabic as Intercept:			
(Intercept)	5.064289	0.068909	73.49
Compounds	0.192431	0.039077	4.92*
Adj-N Phrases	-0.003594	0.032909	-0.11
Disyllabic	0.123905	0.035500	3.49*

Table 7.9: Mean Naming Latencies: Experiment 6

	(1) Compounds	(2) Phrases	(3) Disyll	(4) Mono
PWds	1	2	1	1
LexWds	2	2	1	1
Syllables	2	1	2	1
Mean Lat. (ms)	204 (9.8%)	174 (7.3%)	193 (9.2%)	174 (6.0%)

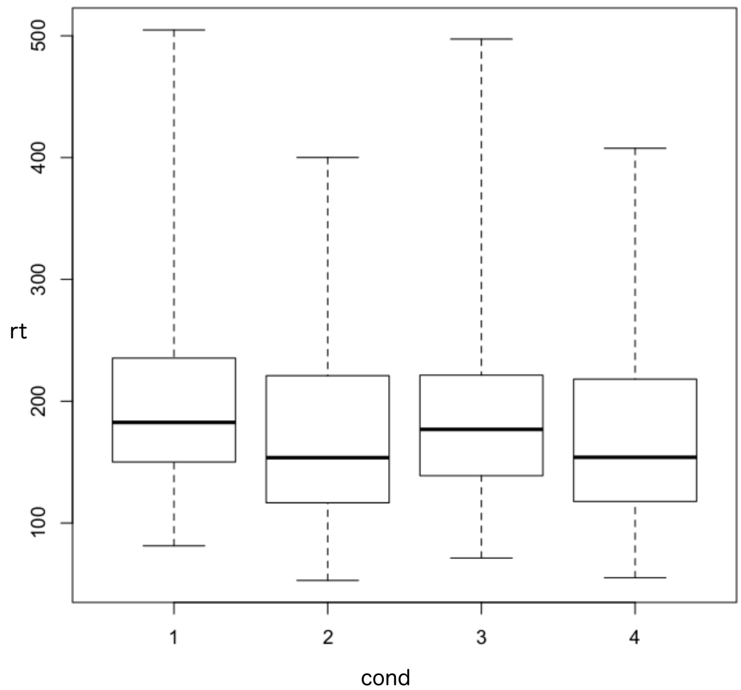


Figure 7.21: Mean Naming Latencies by Condition: Experiment 6

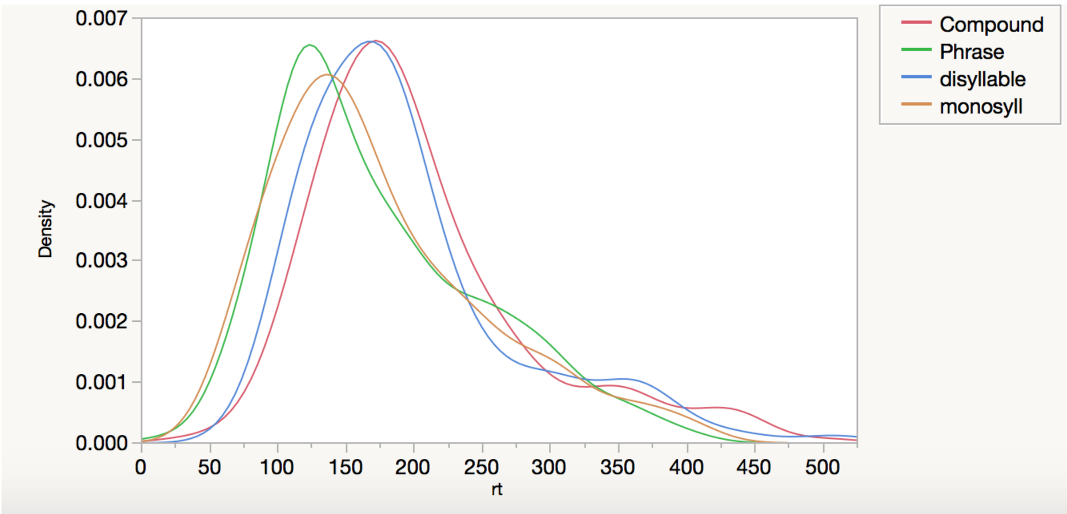


Figure 7.22: Densities of RT by Condition: Experiment 6

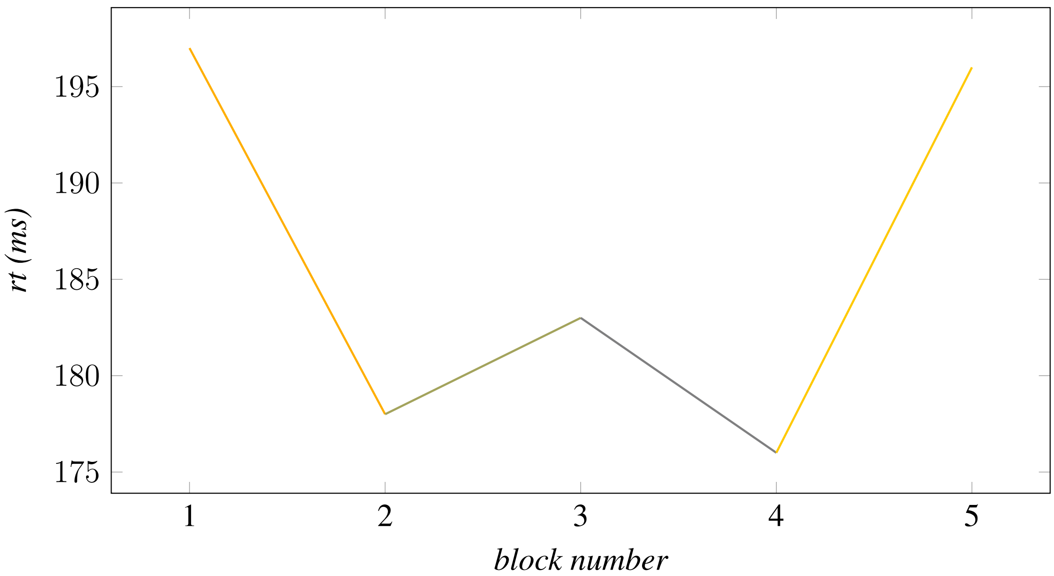


Figure 7.23: Mean RTs by Block: Experiment 6

7.3.4.2 Errors

The error rates were analysed proportionally using a generalised linear model (GLM) with a binomial distribution (link= logit). Disfluencies were the most common error for speakers to make in this task, followed by null responses. The analysis revealed no main effect of condition on error rate (χ^2= 0.45, p= .97). There was a very low number of stress errors due to Condition 4 being monosyllabic (instead of the disyllabic finally-stressed items in Experiment 5). Additionally, there was no effect of block on error rate (χ^2= 7.11, p= .13).

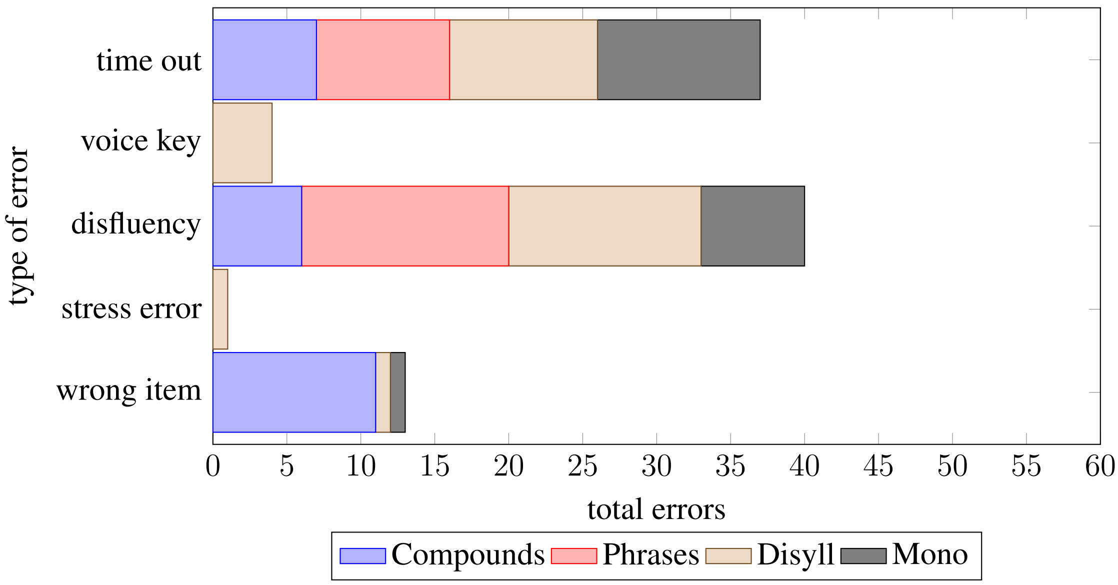


Figure 7.24: Total Errors: Experiment 6

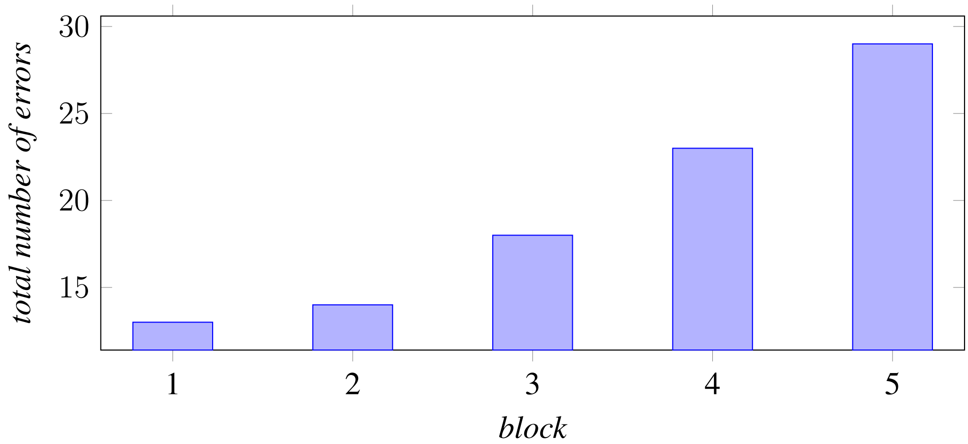


Figure 7.25: Errors by Block: Experiment 6

7.3.4.3 Frequency Analyses

The RT data were again submitted to a linear mixed models analysis, in which frequency and condition were treated as the fixed effect factors. Once again, the maximally-appropriate structure that converged presented random intercepts and slopes for subject factors and a random intercept only for the between items factor.[14]

There was no sign of an interaction between total word frequency and condition in this task: neither with CELEX (Estimate= 0.1129, SE= 0.1481, t= 0.762) nor with log frequency (Estimate= 6.481, SE= 5.042, t= 1.285). Nor was there any evidence of a general effect of word frequency on naming latencies (CELEX: Estimate= -0.1929, SE= 0.1697, t= -1.137; log: Estimate=-6.508, SE= 5.725, t= -1.137). In the compound frequency analysis, neither the first (Estimate= 8.279e-03, SE=1.53202, t= 0.541) nor second (Estimate= 9.85103, SE= 8.48303, t= 1.161), nor the total (Estimate= 0.3117, SE= 0.4593, t=0.679) morpheme frequency reached significance. The frequency of the adjective in the adjective-noun phrase approached significance (Estimate= 0.02562, SE= 0.01293, t= 1.982), while the noun did not (Estimate= -0.006769, SE= 0.006926, t= -0.977). As in Experiment 2, word length elicited no effect (all ts < 1.6) on reaction time in this analysis.

[14]Formula: rt ∼ cond + freq + (1 + cond | sub) + (1 | item).

7.3.5 Discussion

This experiment elicited some very interesting results. First of all, naming latencies were **shorter** than those of native English speakers (compare 186 ms for this task to the 241 ms average for the native speakers). This is contradictory to the delay commonly associated with naming tasks in L2 speakers (cf. Gollan et al., 2005).[15] Nevertheless, these speakers behaved similarly to the native speakers in relation to the planning times for compounds and phrases. Phrases elicited latencies similar to those of the monosyllabic word condition, suggesting that L2 speakers also encoded only the adjective in the phrase before initiating speech. Compounds elicited similar naming latencies to disyllabic words, lending support to our prediction that these highly-fluent L2 speakers would be able to recognise and plan compounds as single prosodic units rather than sequences of units.

Our primary goal in this task was to determine how L2 speakers behaved with the loss of planning time, specifically in relation to the planning of complex morphosyntactic items. We hypothesised that this task would increase the working memory demands on the L2 speaker more than the monolingual speaker: not only would the loss of planning time affect responses, but so would the simultaneous activation of the L1 language. The difficulty of encoding in the L2 was obvious here: the size of the effect in the native task was ∼40 ms, while the non-native task only elicited a difference of ∼15 ms. As expected, these speakers generated more errors overall, a finding that agrees with those of other non-native production studies (cf. Ganushchak & Schiller, 2009).

Errors were different for this task than for the delayed task: instead of "wrong items", the most common errors were disfluencies and "time outs". There were very few stress errors due to the removal of the finally-stressed disyllabic targets in Condition 4. Overall, speakers either stuttered when initiating speech, or said nothing. This also reflects the difficulty associated with online naming in the L2, and in particular the increased

[15]The vast majority of evidence of this phenomenon comes from L2 naming tasks involving picture naming. Picture-naming tasks require activation of appropriate lemmas, something that is already taken care of by the visual prompts in these studies.

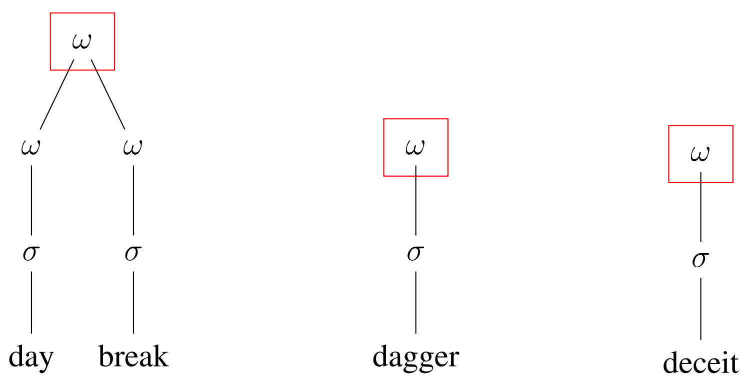
demands on working memory to speak as quickly as possible. As with the monolingual task in Experiment 2, word length had no effect in this task. This lends strength to the articulatory approach to single-word naming discussed in Damian et al. (2010). This is another example of the L2 speakers behaving like the L1 speakers.

7.4 General Discussion

The aim of these experiments was to elicit psycholinguistic data about the generation of prosodic frames in non-native speakers of English. Our hypotheses for both experiments hinged upon the speakers' fluency levels: we predicted, on the basis of existing data, that speakers would be able to access the correct prosodic frames for compounds and phrases if they were fluent enough in English. Following this, we also theorised that these tasks would exhibit certain 'markers' of L2 naming studies: an increase in errors, an increase in overall latencies, and a decrease in effect size.

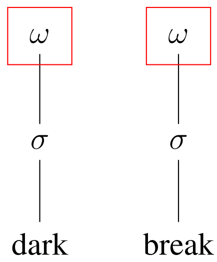
When given time to plan their utterances, highly-fluent L2 English speakers behaved as expected: they treated noun-noun compounds as single prosodic units.

(7.2) Units predicting naming latencies in compounds and simple words



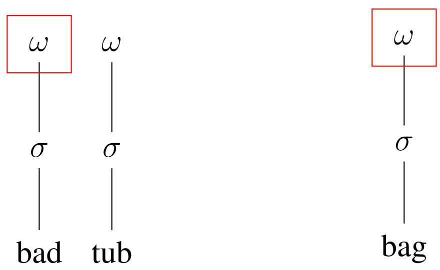
Utterances containing adjective-noun phrases elicited significantly-longer naming latencies.

(7.3) Units predicting naming latencies in adjective-noun phrases

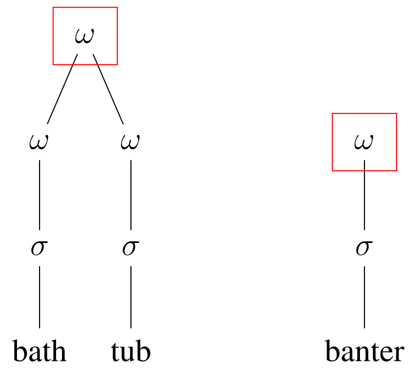


Both of these results were in line with those found for native English speakers. When planning time was taken away in the online task, the amount that L2 speakers were able to prepare before initiating articulation changed; reaction times for this experiment reflected the size of the first prosodic unit of the utterance. For compounds, this was a single, disyllabic Phonological Word (e.g. *bathtub*); for phrases, it was a monosyllabic Phonological Word (e.g. only the *bad* in *bad tub*).

(7.4) Restricted planning scope in Experiment 2: Phrasal and monosyllabic word conditions



(7.5) Planning of compounds and disyllabic words in online task conditions



This was also similar to how the native English speakers planned their speech in online task conditions. The size of main effect reflected the increased task load of speaking in a non-native language (cf. Ando et al., 2014): in the delayed task, there was only an average 15 ms difference between compounds and phrases for L2 speakers, in comparison to an average 40 ms difference for native speakers.

We have evidence in these experiments that the planning scope is also flexible for L2 speakers, and that it is dependent upon fluency in the L2. If the speaker is able to access the correct morphemes, phonemes, and prosodic frame, then they generate results similar to L1 speakers. If they are not (in the case of Experiment 5b), then they generate flat results.

Errors revealed interesting details about the non-native production process: aside from stress errors, the most common error the highly-fluent group made was "incorrect item". In contrast, the most common error for the low fluency group was a "time out". Where highly-fluent speakers often replaced target words with other words containing (at the very least) the same initial phoneme (e.g. *bell* for *ballad*), the less-fluent speakers simply said nothing.

We saw a very large number of stress errors associated with the finally-stressed control condition (Condition 4) in Experiment 5. This was entirely predictable due to the Bengali stress rule, which stipulates that stress in Bengali is overwhelmingly word-initial. It is notable that, in most cases, the subjects produced the correct word, only with an incorrect stress pattern (*gázèlle* for *gàzélle*. In the instances where speakers are able to assign correct stress, overruling the default stress of the L1, but this raises the working memory demands for the task even more, resulting in an increase in errors and longer reaction times- both of which we saw in Experiment 5.

Results from a number of naming tasks have indicated a priming effect of L2 words that share phonemic representations with the corresponding L1 words and, in particular, share initial phonemic segments (e.g. Eng. *month* and Fr. *mois*). A priming effect has also been seen in naming tasks involving cognate words (e.g. Eng. *cold* and Gr. *kalt*) (Costa et al., 2000; Gollan & Acenas, 2004). However, Roelofs & Verhoef (2006) claim that the sharing stops here: as evidenced from our results for the disyllabic final-stressed targets of Experiment 5, rules of stress assignment are certainly not shared between English and Bengali. Rules of syllabification are also language-specific; evidence for this comes from Cutler et al. (1986), who found that native French speakers defaulted to regular French syllabification even when processing English words. Our results support this claim; when the non-native speakers were not familiar with a word enough to force the iambic English stress pattern onto it, they defaulted to their trochaic Bengali stress patterns.

Now we have a starting point for the processing of spoken English in non-native speakers. We know that fluency is an important factor in these experiments: unless speakers are able to identify both the targets and their intended prosodic shapes, they will not be able to generate the correct prosodic frames for the words. This will become crucial as we increase the task complexity in the next set of experiments: if speakers are not fluent enough to identify compounds as single prosodic units, they may not be able to plan the clitic conditions well.

These experiments have also given us a rare look at online language production in L2 speakers: our non-native subjects were actually faster in the online task than the native speakers. This may be due to a number of factors, including 'over-rehearsing' from the L1 speakers. However, this speed came at a cost; errors were noticeably higher for the non-native speakers in the online task. In summary, we have evidence that L2 speakers, with enough fluency, generate English prosodic frames much in the same way as native English speakers, albeit at a price.

8 The Phonological Encoding of Clitics by Non-Native English Speakers

8.1 Introduction

In Experiments 3 and 4, we saw that native English speakers planned structures containing both compounds and clitics as single prosodic units:

(8.1) (a) (groundhogs are)$_{\omega}$ (nice)$_{\omega}$

(b) (doorways are)$_{\omega}$ (clean)$_{\omega}$

Naming latencies for these items reflected that the auxiliary *are* reduced phonologically and enclitised leftwards onto the compounds as a single unit. This resulted in similar naming latencies to both disyllabic and monosyllabic targets that included clitics:

(8.2) (a) (dolphins are)$_{\omega}$ (nice)$_{\omega}$

(b) (drapes are)$_{\omega}$ (clean)$_{\omega}$

The experiments in this chapter test the planning of these structures in L2 speakers of English. Recall the results of Experiments 5 and 6 in the previous chapter: we saw that planning at the phonological encoding level was affected not only by the preparation time, but also by fluency level. The less-fluent speakers (cf. §8.2.4) made more errors, had longer naming latencies and, most importantly for this study, were not able to distinguish between the experimental conditions. Given these results, this chapter investigates the planning of structures containing compounds and clitics in highly-fluent non-native English speakers.

The outline of this chapter will remain the same as in previous experimental chapters: first, we test the planning of complex items in a delayed production task. §8.2.2 offers several hypotheses based on the information we have gathered thus far. Results of the reaction time, error, and frequency analyses can be found in §8.2.4, followed by a discussion and interim summary. The second part of the chapter contains the final experiment, an online production task designed to measure the naming latencies of compounds and clitics in non-native English speakers. §8.3.1 presents hypotheses, §8.3.3 presents results, and §8.3.4 contains a discussion of the results. Finally, the chapter concludes with a general discussion about the overall implications of the results of these experiments in §8.4.

8.2 Experiment 7: Delayed Production of English Clitics in Native Bengali Speakers

The experiment reported in this section was designed to elicit information about how non-native English speakers plan cliticised compounds when they have time to prepare their utterances. This experiment consists of a delayed production task with a question-answer paradigm, identical to that of Experiment 3.

8.2.1 Hypotheses

The primary aim of this experiment is to increase the complexity of the prosodic unit and elicit information for how it is planned during L2 phonological encoding. In light of the results of Experiment 3 and 4, as well as existing empirical evidence about encliticisation, the main hypotheses tested in this study are as follows:

1. Despite an increase in task complexity, the naming latencies of the fluent non-native speakers will continue to reflect the number of prosodic, not lexical, units.
Results from Experiments 5 and 6 implied that L2 speakers of English exhibited a flexible planning scope, as long as they were fluent enough to complete the task. Crucially, while the planning scope varied depending on the task conditions (i.e. delayed or online), the unit planned by non-native speakers was the same as in native English speakers: the Phonological Word.

This experiment introduces another layer of complexity for the participants: Recall that this task (reproduced from Experiment 3) requires speakers to answer one of five questions:

Table 8.1: Auditory Prompts: Experiments 3, 4, 7, 8

1. What are good?
2. What are nice?
3. What are clean?
4. What are dry?
5. What are big?

This was done in order to encourage encliticisation of the auxiliary verb to the target word stimuli. In native speakers, we saw an effect of word length in the delayed clitic production task: there was a difference of 60 ms between the longest and shortest words. Preparation time was also a factor for native speakers in this task: the shortest preparation time resulted in the longest naming latencies. Following these results, it seems likely that the L2 speakers will exhibit signs of task demands. Of all the experiments, these will be the most complex: not only will L2 speakers be required to plan multiple recursive structures (in the case of compounds) in their non-native language, but they will also be required to answer with the correct prompt.

In delayed task conditions (Experiments 1, 3, and 5), we have found that it is the total number of Phonological Words in the utterance that predicts naming latency. Despite an increase of task complexity, we predict that the speakers in this task will have sufficient proficiency to continue planning well-formed prosodic units in English at during the phonological encoding stage. That is, the frames into which the segments are arranged will continue to reflect well-formed Phonological Words. Related to this level of fluency, we also predict that L2 speakers will treat compounds no differently than they did in Experiments 5 and 6: as single, recursive phonological units. This brings us to our next hypothesis.

2. Speakers will continue to treat compounds as single prosodic units, and the auxiliary *are* will exhibit vowel reduction and encliticisation to the compound to form a super-sized recursive Phonological Word.

The phenomenon of encliticisation is a prominent feature of spoken English, and the reduction of vowels in auxiliaries such as *are* is also very regular (cf. Labov, 1969). Moreover, there is ample evidence that L2 speakers, particularly those of less fluency, regularly fail to inflect verbs correctly in English (Meisel, 1997; Lardier, 1998): this phenomenon has been attributed to either an absence of inflectional features in the speaker's mental grammar, or problems related to encoding accurate morphological representations (Ionan & Wexler, 2002). Regarding the reduction and encliticisation of auxiliaries in L2 speech, there is evidence that L2 speakers of English behave differently to L1 speakers (cf. Baker et al, 2011). Phonetic analyses of non-native speech patterns have revealed that L2 speakers of English on average produce less syllable reduction (Anderson-Hsieh & Venkatagiri, 1994) and less function word reduction (Aoyama & Guoin, 2007).[1] Ueyama (2000) elicited evidence that advanced Japanese learners of

[1]This phenomenon is not restricted to English: in a study investigating vowel reduction in L2 speakers of German, Gut 2003 found that L2 speakers tended to reduce auxiliary and function words less than L1 speakers.

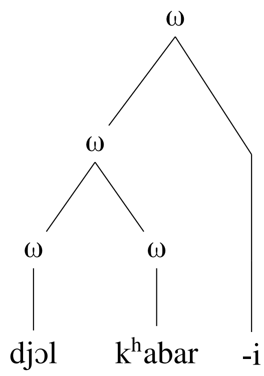
English exhibited more frequent durational reduction of unstressed syllables in function words than beginning learners did, but on average L2 Englsih speakers exhibited less reduction overall than L1 speakers (61).[2] From these results, it appears that fluency plays a role in how L2 speakers treat auxiliary verbs. Ueyama proposed that the overall effect of L2 reduction may be related to transfer from the native language: in her study, the speakers had Japanese as their L1, a language in which independent function words are absent (Baker et al, 2011: 4). This leads us to also consider our speakers' native language.

Fitzpatrick Cole (1990) addressed cliticising and compounding in Bengali in her study on focus intonation: in her view, the Phonological Word in Bengali behaves much like its English counterpart, particularly in respect to clitic attachment. She assigns phonological processes such as vowel lengthening to the Phonological Word level, arguing that they are dependent on well-formed phonological units. Bengali compounds and clitics rely upon well-formed Phonological Words: clitics attach to a Phonological Word, and compounds are formed of two or more Phonological Words.

Cliticisation is productive in Bengali, and clitics often attach to compound words just as in English: for example, the emphatic clitic [=i] in the compound *djɔlk^{h}abari* ("snack!"):

8.1 *djɔlk^{h}abar= i* *tʃai*
water-food EMPH want(1P)
It's definitely a snack that I want.

[2]Ueyama's reasoning behind this was that L2 speakers "tend to import L1 phonetic habits in L2 word accent production" (62). However, as we saw in the results from the native Bengali speakers of the disyllabic final-stressed word condition (§7.2.4), the influences of L1 features on L2 speech extend far beyond "phonetic habits": the default trochaic Bengali stress pattern resulted in errors such as *gázèlle* for *gàzélle* and *lápel* for *làpél*.



"water-food (snack)"

In the above example, the emphatic clitic /=i/ attaches to the noun *djɔlkʰabar*, which stands as a single phonological unit.

Based on the data elicited from Experiments 5 and 6, as well as our understanding of Bengali Phonological Word structure, we theorise that highly-fluent Bengali native speakers will continue to treat compounds as single prosodic words, and that the auxiliary verb *are* will reduce and cliticise to this unit. Once again fluency will be key in this analysis: the more fluent the speaker is, the higher probability they have of producing this intended structure.

8.2.2 Stimuli Selection

Table 8.2: Stimuli: Experiment 7

Condition 1: Compounds	**Condition 2: Phrases**	**Condition 3: Disyllabic**	**Condition 4: Monosyllabic**
doorways	deep ways	daisies	doves
dishcloths	drab cloths	dolphins	ducks
dustpans	dark pans	donkeys	drapes
graveyards	green yards	gospels	grapes
grandstands	grey stands	griddles	graphs
grindstones	good stones	goblins	gloves
nightgowns	nice gowns	noodles	nuns
neckties	neat ties	napkins	nets
nightshirts	new shirts	nickels	nodes
lampshades	low shades	lemons	leeks
logbooks	late books	lanterns	lungs
lipsticks	large sticks	lions	lanes
bookshops	big shops	barrels	bowls
bathtubs	bright tubs	blankets	blades
ballrooms	blue rooms	bankers	brooms

8.2.3 Participants and Procedure

Twenty participants between the ages of 14 and 21 took part in Experiment 7. They were recruited from Shri Shikshayatan School in Kolkata, India, in March 2015. Participants were native speakers of Standard Bengali and had attended an English-language track school since age 5. These speakers all scored very high on the proficiency tasks; no participant scored lower than Band 8 in the speaking task, and most were comfortably within Band 8 or 9 (see Appendix C for the band descriptors for this task). They were given course credit for their participation. Due to equipment constraints, the experimenter was seated at the far end of the same room, facing away from the participant. The rest of the design and procedure were identical to Experiment 3.

8.2.4 Results

8.2.4.1 Latencies Analysis

Analyses were identical to Experiment 3. Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error. Utterances which did not exhibit reduction and cliticisation were also marked as errors: this was done by measuring the duration of the vowel in the auxiliary verb. Responses uttered before the final beep were also discarded. All data points beyond two standard deviations from the mean were counted as outliers and removed. This resulted in a loss of 13.2% of data.

Following Bates et al. (2014), we began by testing the maximally-appropriate random structure and continued to adjust the model until it reached convergence. The maximal model contained a full interaction between condition and preparation time in the fixed effects, an interaction with random slopes between condition and beep in the by-subject analysis, and intercepts for condition and beep in a by-item analysis. This model[3] failed to converge. The next maximal model to converge had an interaction between condition and beep in the fixed effect, random slopes and intercepts for condition in the by-subject analysis, and intercepts for condition in a by-item analysis.[4] This model converged. The output of this model is shown in Table 8.3. This scatterplot for this model had some striping in it; however the histogram and QQ-plots appeared homoscedastic, as shown in figure 8.1. Crucially, a log-transform did not improve model fit.

Following Baayen (2008), all t-values greater than 2 were treated as significant. The results from the mixed-effect analysis of condition on reaction time revealed a significant difference in the naming latencies for adjective-noun phrases (t= 4.148*). In order to generate pairwise comparisons between the four conditions, each condition type was

[3] mla.maxstar <- lmer(rt ∼ cond*beep + (1+cond*beep|sub)+ (cond+beep|item), data=d, REML=FALSE).
[4] (rt ∼ cond+beep+ (1+cond|sub)+ (1|item),data=d)

systematically treated as the intercept. Adjective-noun phrases differed significantly from the other three conditions, while noun-noun compounds showed no significant difference to either of the morphologically-simple word conditions. These analyses also showed that simple words differed significantly from phrases, but not from one another.

Mean reaction times and percentage error rates for Experiment 7 are shown in Table 8.4. The graph below (Figure 8.2) shows the distribution of mean naming latency (RT) for all four conditions. Preparation time (beep) was significant in this experiment, eliciting significantly different naming latencies for each preparation time (Table 8.5). There was no interaction between preparation time and condition. The results also revealed an effect of block on RT (Figure 8.4), in which targets became significantly slower in the second block (t=-2.220*), but once again there was no interaction between block and condition.

Table 8.3: Linear mixed-effects analyses of RT data in Experiment 7

	Estimate	**SE**	**t score**
All conditions:			
(Intercept)	429.334	25.569	16.791
Adj-N Phrases	23.656	5.704	4.148*
Disyllabic	-1.612	4.580	-0.352
Monosyllabic	5.245	4.850	1.081
Pairwise Comparisons:			
Adj-N Phrases as Intercept:			
(Intercept)	452.990	26.518	17.082
Compounds	-23.656	5.704	-4.148*
Disyllabic	-25.268	5.500	-4.594*
Mono	-18.412	6.024	-3.056*
Disyllabic as intercept:			
(Intercept)	427.722	26.025	16.435
Compounds	1.612	4.580	0.352
Adj-N Phrases	25.268	5.500	4.594*
Monosyllabic	6.856	4.531	1.513
Mono as Intercept:			
(Intercept)	434.578	26.823	16.201
Compounds	-5.245	4.850	-1.081
Adj-N Phrases	18.412	6.024	3.056*
Disyllabic	-6.856	4.531	-1.513

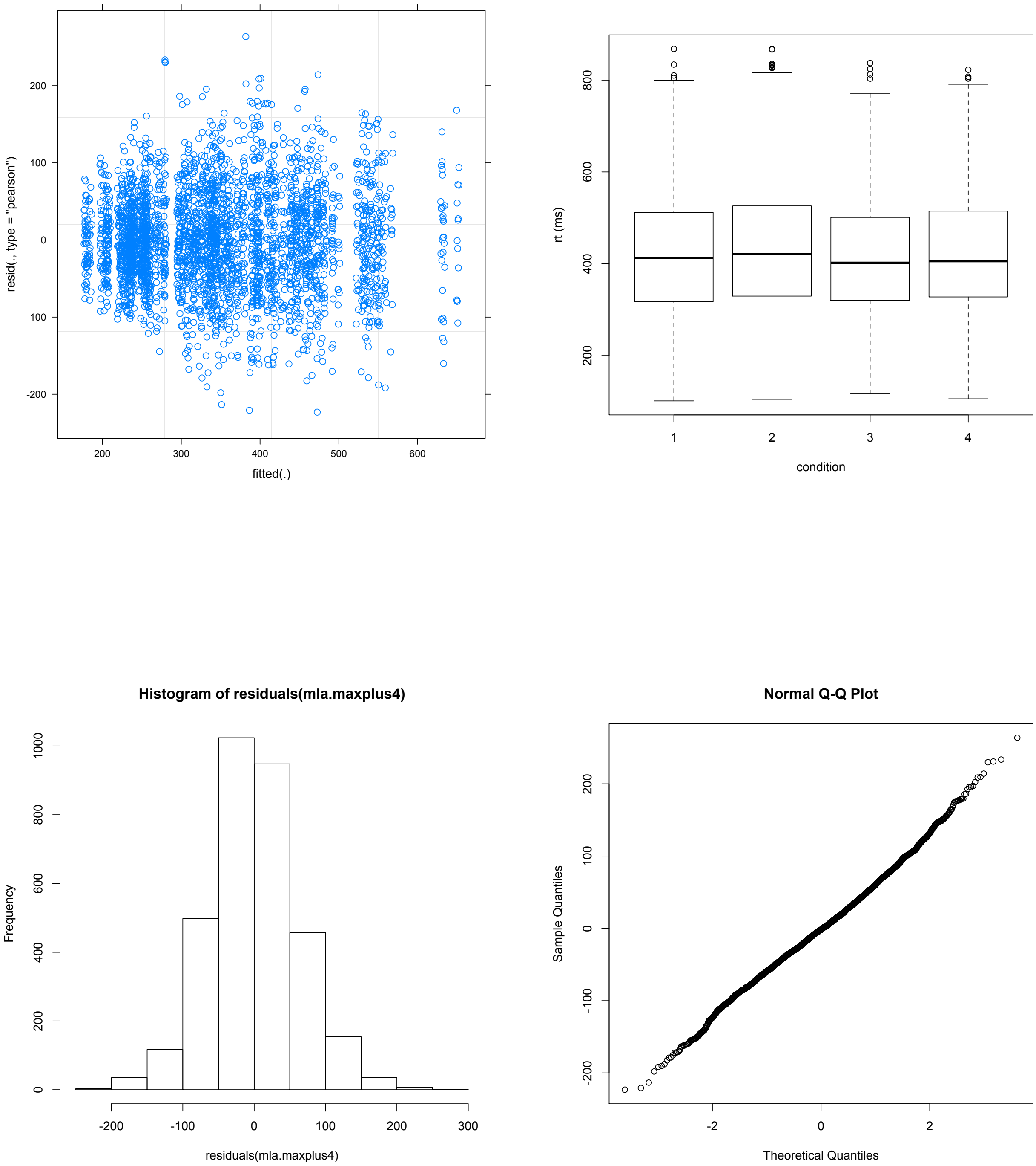


Figure 8.1: Residuals Analyses: Experiment 7

Table 8.4: Mean Naming Latencies (in ms): Experiment 7

	(1) Compounds +Clitics	(2) Phrases +Clitics	(3) Disyll +Clitics	(4) Mono +Clitics	Mean Lat. (ms)
PWds	1	2	1	1	
LexWs	2	2	1	1	
Syllables	2	2	2	2	
Beep Lat.:					
800 ms	440 (%)	453 (%)	425 (%)	436 (%)	439 (8.2%)
1200 ms	409 (%)	433 (%)	415 (%)	414 (%)	418 (7.7%)
1400 ms	400 (%)	427 (%)	403 (%)	406 (%)	409 (5.3%)
Mean Lat. (ms)	416 (8.5%)	438 (5.0%)	414 (6.2%)	419 (8.1%)	

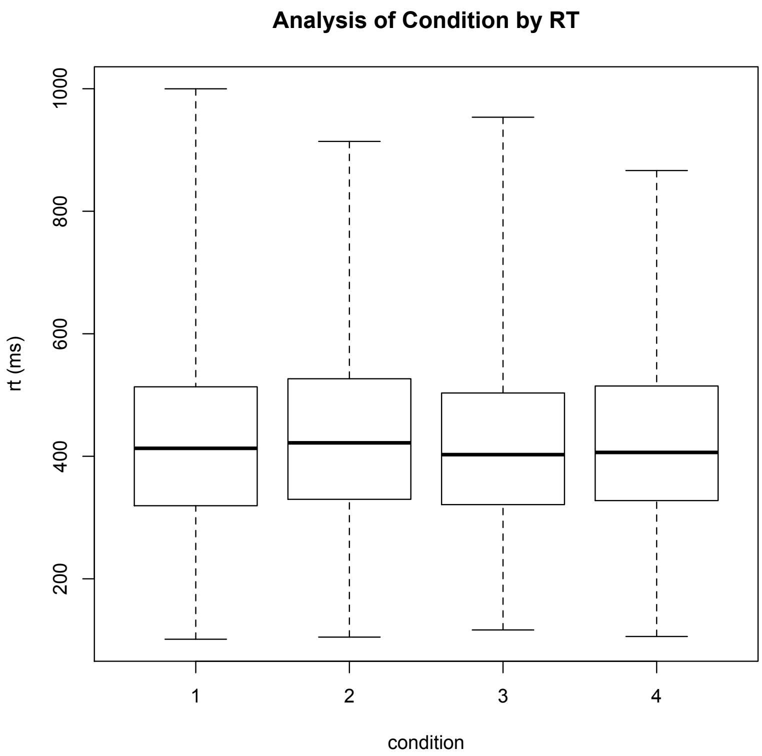


Figure 8.2: Mean Naming Latencies by Condition: Experiment 7

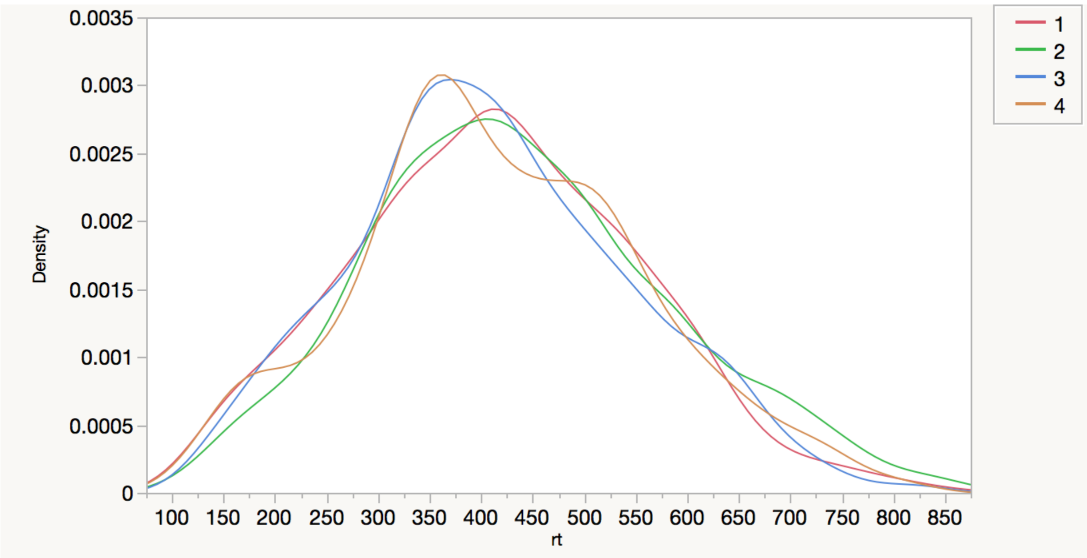


Figure 8.3: Densities of RT by Condition: Experiment 7

Table 8.5: Linear Mixed-Effects Analyses of Beep*RT Data: Experiment 7

	Estimate	**SE**	**t/z score**
Beep1 (800 ms) as intercept:			
(Intercept)	429.334	25.569	16.791
1200 ms	-19.136	3.398	-5.631*
1400 ms	-27.084	3.398	-7.970*
Beep2 (1200 ms) as intercept:			
(Intercept)	410.197	25.566	16.045
800 ms	19.136	3.398	5.631*
1400 ms	-7.948	3.379	-2.352*
Beep3 (1400 ms) as intercept:			
(Intercept)	402.250	25.566	15.734
800 ms	27.084	3.398	7.970*
1200 ms	7.948	3.379	2.352*

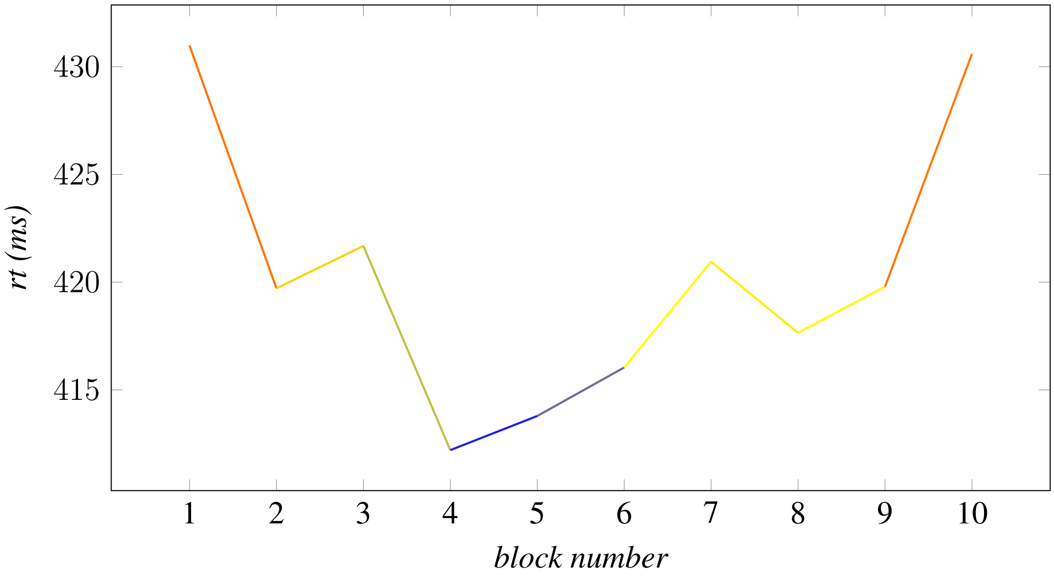


Figure 8.4: Mean RT by Block: Experiment 7

8.2.4.2 Errors

The error rates were analysed proportionally using a generalised linear model (GLM) with a binomial distribution (link= logit). While these speakers did not make frequent errors, disfluency errors were the most common, followed by wrong items and time outs (see Figure 8.5 below). The analysis revealed no main effect of condition on error rate (χ^2= 6.6, p= 0.15). Preparation time was revealed to have an effect on error rates (χ^2= 8.51, p= 0.014*): the shortest preparation time (800 ms) resulted in the highest number errors (χ^2= 8.41, p= .0037*). There were no other significant effects.

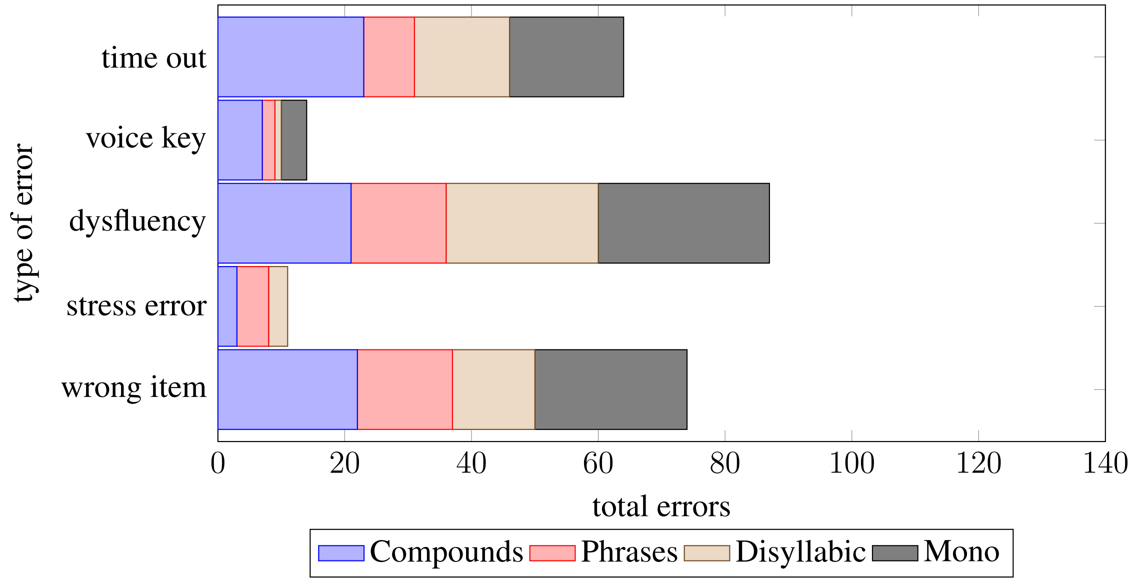


Figure 8.5: Total Errors: Experiment 7

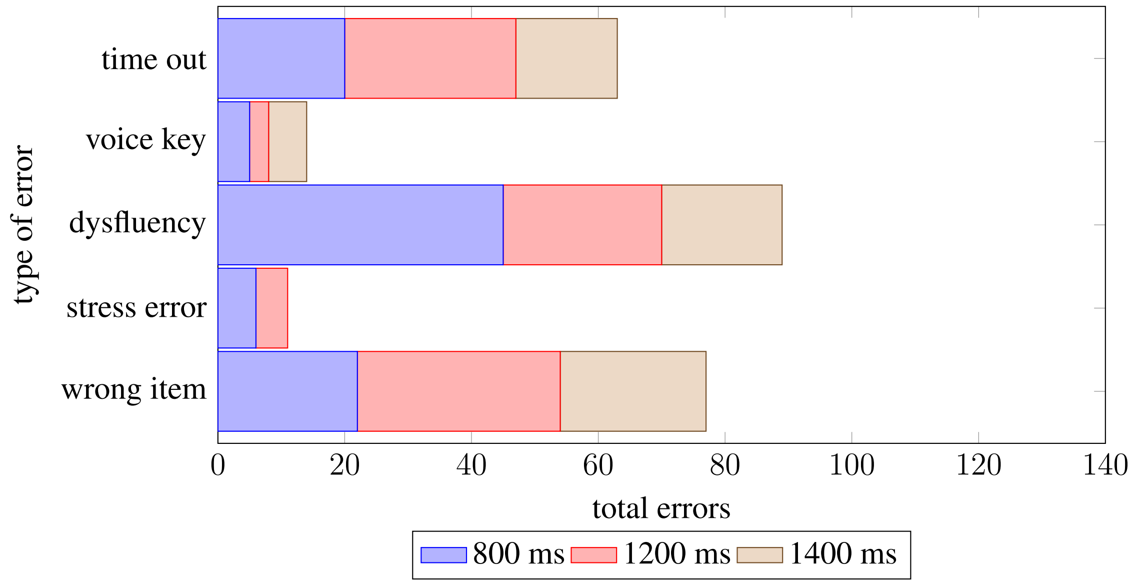


Figure 8.6: Error by Preperation Time: Experiment 7

8.2.4.3 Frequency Analyses

The maximally-appropriate structure that converged presented random intercepts and slopes for subject factors and a random intercept only for the by-items factor.[5] Neither log (Estimate =1.114, SE= 4.430, t= 0.251) nor CELEX (Estimate= 0.01834, SE= 0.21886, t= 0.084) measures of total word frequency had an effect on naming latencies in this task. In

[5]Formula: rt ∼ cond + freq + beep + (1 + cond | sub) + (1 | item).

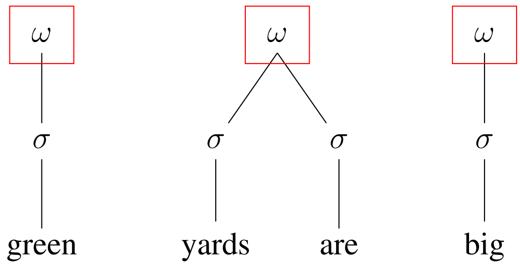
the compound frequency analysis, neither the first (Estimate= -0.03654, SE= 0.02170,t= -1.684), second (Estimate= -0.01236, SE= 0.01139, t= -1.085), nor total (Estimate= -2.119, SE= 6.566, t= -0.323) morpheme word frequency affected naming latency. Neither the adjective morpheme frequency (Estimate= -0.006138, SE= 0.010376, t= -0.592) nor the noun (Estimate= -0.008458, SE= 0.01029, t= -0.822) morpheme frequency affected naming latencies. Lastly, an analysis of the effect of word length (number of letters) did not result in any significant values (all ts < 1.3) for any of the lengths on naming latency.

8.2.5 Discussion

This experiment elicited a significant main effect on naming latencies across all four conditions. Speakers took, on average, 22 ms longer to prepare the phrasal + clitic condition than they did all other conditions. This suggests that compounds are again being treated as single Phonological Words, while phrases are being treated as two distinct Phonological Words. A cross-analysis of the naming latencies of compounds with both types of morphosyntactically-simple words did not result in any significant difference, lending evidence to the hypothesis that these three items are being treated similarly by speakers.

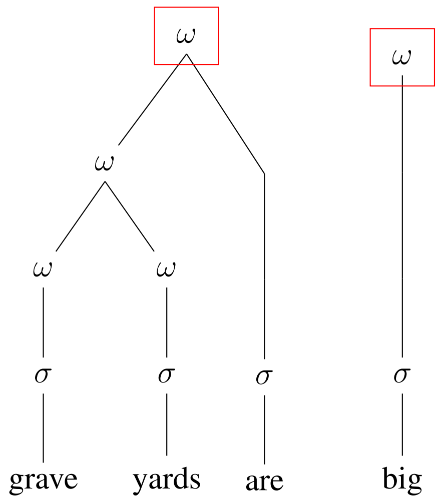
Returning to our predictions, we find evidence that speakers were indeed able to plan speech in prosodic, not lexical, units. Despite the increase of task demands, latencies continued to reflect the planning of prosodic units, not lexical units. For adjective noun phrases, the target items consisted of two prosodic units: the adjective, and the noun + clitic PWd structure (the total utterance contains three prosodic units) (example 8.1):

(8.1) Units predicting naming latencies in adjective-noun phrases



We also found that the compounds behave as they did in the native speakers: as single prosodic units. Although compounds are structurally much more complex (see Example 8.2) than either morphosyntactically-simple word conditions (Example 8.3), the L2 speakers did not treat them any differently in terms of naming latencies.

(8.2) Units predicting naming latencies in compounds



(8.3) Units predicting naming latencies in morphosyntactically-simple words

The representations in the examples above show the phonological units (in red) that are generating the naming latencies for the compound + clitic condition, and morphosyntactically-simple word + clitic conditions. In the compound representation, it is the recursive Phonological Word unit of the compound as a whole to which the clitic attaches.

These results suggest that, when given time, L2 speakers were able to generate the correct prosodic frames. The introduction of the variable auditory prompt did not generate a significantly-higher number of errors than for the L1 speakers. Additionally, there was no "illegal" condition (as in Experiment 5's disyllabic word condition containing iambic stress), so there were no incorrect frame activations. The mean reaction times per block showed that as speakers settled into the task, they began to reply faster and more confidently towards the middle of the experiment.

Contrary to the results for the L2 speakers in Experiment 5, we did not elicit any frequency effects in this experiment. Nor did we find an effect of word length on reaction times. However the task difficulty was visible in other areas: all three preparation times had significantly different naming latencies, with the shortest preparation time (800 ms) generating latencies 18 ms slower than the 1400 ms deadline.

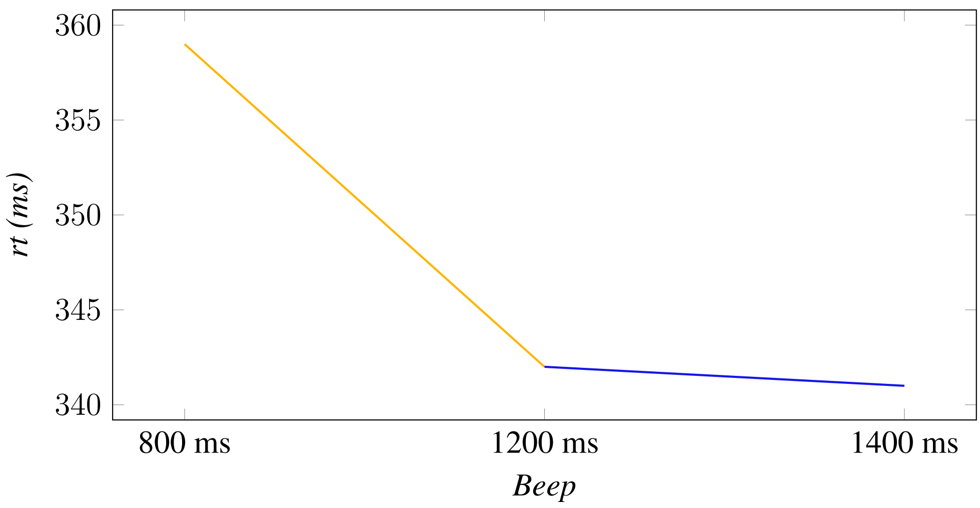


Figure 8.7: Mean RT by Preparation Time: Experiment 7

Errors were highest for the compound + clitic conditions and lowest with the phrasal conditions. As with previous results, there was no effect of frequency or word length on naming latencies. Errors for this task were not particularly high: disfluencies, time outs and wrong items were the most common types made. The shortest preparation time resulted in the highest number of errors.

However, while this task shows that L2 speakers are behaving like the L1 speakers, it does not necessarily present solid evidence that clitics are reducing and attaching to the neighbouring word. The difference in naming latencies between compounds and phrases could reflect that speakers are treating the auxiliaries as independent prosodic units. Furthermore, it does not show which direction the clitics are attaching. The only way to do this is to test the size of the first planning unit: luckily, the online task does exactly that.

8.3 Experiment 8: Online Production of English Clitics in Native Bengali Speakers

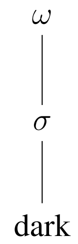
8.3.1 Hypotheses

In Experiment 7, we gained evidence that, despite being non-native English speakers, the Bengali speakers were treating compound words as single prosodic units. Following these results as well as keeping in mind the results of the online task in Experiment 4, we propose the following predictions below:
The method, materials, and procedure are identical to Experiment 4.

1. **The loss of planning time will affect the size of the planning unit as well as the process that prepares it for articulation.**
In both the L1 and L2 online tasks presented in this study, we have seen a significant effect of the removal of planning time. In all tasks, the results suggest that speakers are only able to access and plan the first phonological unit before initiating utterance. Therefore, we maintain that the planning scope of both native and non-native speakers exhibits flexibility depending on task type and target complexity.

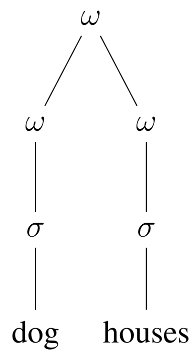
Of all four experiments conducted thus far, this will be the most difficult for the non-native speakers. Not only will they be required to respond to the prompt correctly, they also must answer as quickly as possible. The combined difficulty of this task may generate more errors and deviations than we have previously seen in online tasks. However, following the results of Experiment 4, we hypothesise that the L2 speakers will attempt to plan the correct (English) prosodic frames. That is, they will continue to treat phrases as as two discrete phonological units, and thus will only access the first Phonological Word (the adjective) of the condition:

(8.4) dark



Conversely, compound words will elicit significantly longer naming latency, due to their size:

(8.5) Doghouses are nice.



Following this, we predict that clitics will attach to the higher (outer) prosodic word of the compounds. This leads us to our next, and final, hypothesis:

2. **The auxiliary will reduce and attach leftwards.**

Results from Experiment 7 support our hypothesis that L2 speakers are still be able to plan compounds as single prosodic units, and therefore, we propose that the auxiliary *are* will reduce and attach to the compound as a single unit. Naming latencies for the compound and simple morphosyntactic words (both monosyllabic and disyllabic) were statistically similar in Experiment 7, lending some support to the prediction that these items are treated the same during prosodification. We believe that the naming latencies in Experiment 7 reflect a recursive Phonological Word structure in the compound condition and therefore, we will see evidence for the size of this structure in the naming latencies for this task.

8.3.2 Participants

Twenty-five participants between the ages of 18 and 22 took part in Experiment 8. Participants were native speakers of Standard Bengali. They were recruited from Shri Shikshayatan School in Kolkata, India, in March 2015. Participants were native speakers of Standard Bengali and had attended an English-language track school since age 5. These speakers all scored very high on the proficiency tasks; no participant scored lower than Band 8 in the speaking task, and most were comfortably within Band 8 or 9 (see Appendix C for the band descriptors for this task). They were given course credit for their participation. Due to equipment constraints, the experimenter was seated at the far end of the same room, facing away from the participant. The rest of the design and procedure were identical to Experiment 4.

8.3.3 Results

8.3.3.1 Latencies Analysis

Responses that contained disfluencies, null responses (the subject said nothing), or incorrect answers were discarded from this analysis by an impartial coder. Any difference between the intended sentence and the produced sentence in lexical or syntactic structure was marked as an error. The total number of errors was low, with only 5.9% of data being discarded due to these parameters. Data trimming treated all data points beyond two standard deviations from the mean as outliers: this resulted in the removal of an additional 102 data points. The total loss made up 9.1% of data.

Following Bates et al. (2014), we began by testing the maximally-appropriate random structure and continued to adjust the model until it reached convergence. The maximal model contained a full interaction with condition in the fixed effects, an interaction with random slopes between condition in the by-subject analysis, and intercepts for condition in a by-item analysis. This model[6] failed to converge. The next maximal model to

[6] mla.maxstar <- lmer(rt ~ cond + (1+cond|sub)+ (cond|item), data=d, REML=FALSE).

converge contained an interaction in the fixed effect factors, a random slope for the by-subject analysis, and an intercept for the by-item analysis.[7] This model converged and appeared fairly homoscedastic when fitted to residual plots, shown in figure 8.8. The Q-Q plot showed some evidence of bending; however the histogram was fairly evenly distributed. Crucially, a log-transform did not improve model fit.

Following Baayen (2008), all t-values greater than 2 were treated as significant. The results from the mixed-effect analysis of condition on reaction time revealed a significant difference in the naming latencies for adjective-noun phrases (Estimate= -64.085, SE= 6.094, t= -10.516*). In order to generate pairwise comparisons between the four conditions, each condition type was systematically treated as the intercept. Adjective-noun phrases differed significantly from the other three conditions, while noun-noun compounds showed no significant difference to the disyllabic word condition. These analyses also showed that simple words differed significantly from one another. These values can be seen in Table 8.6, where significant effects are marked with an asterisk. Mean reaction times and percentage error rates for Experiment 8 are shown in Table 8.7. The graph below (Figure 8.9) showed the distribution of mean naming latency (RT) for all four conditions. The results also revealed an effect of block (Figure 8.11), in which targets became significantly faster in the second block and remained so (Estimate= -13.128, SE= 5.037, t= -2.606*).

[7] mla.maxplus3 <- lmer(rt ∼ cond + (1+cond|sub)+ (1|item))

Table 8.6: Linear Mixed-Effects Analyses of RT Data in Experiment 8

	Estimate	**SE**	**t score**
All Conditions:			
(Intercept)	305.581	17.852	17.117
phrases	-64.085	6.094	-10.516*
disyllabic	-5.625	4.591	-1.225
mono	-37.606	5.764	-6.525*
Pairwise Comparisons:			
Phrases as Intercept:			
(Intercept)	241.233	17.340	13.912
Compounds	65.603	4.161	15.767*
Disyllabic	59.539	4.113	14.477*
Mono	27.053	4.123	6.562*
Disyllabic as Intercept:			
(Intercept)	300.772	17.333	17.352
compounds	6.063	4.133	1.467
phrases	-59.539	4.113	-14.477*
mono	-32.487	4.092	-7.938*
Mono as Intercept:			
(Intercept)	268.285	17.335	15.476
compounds	38.550	4.141	9.309*
phrases	-27.053	4.123	-6.562*
disyllabic	32.487	4.092	7.938*

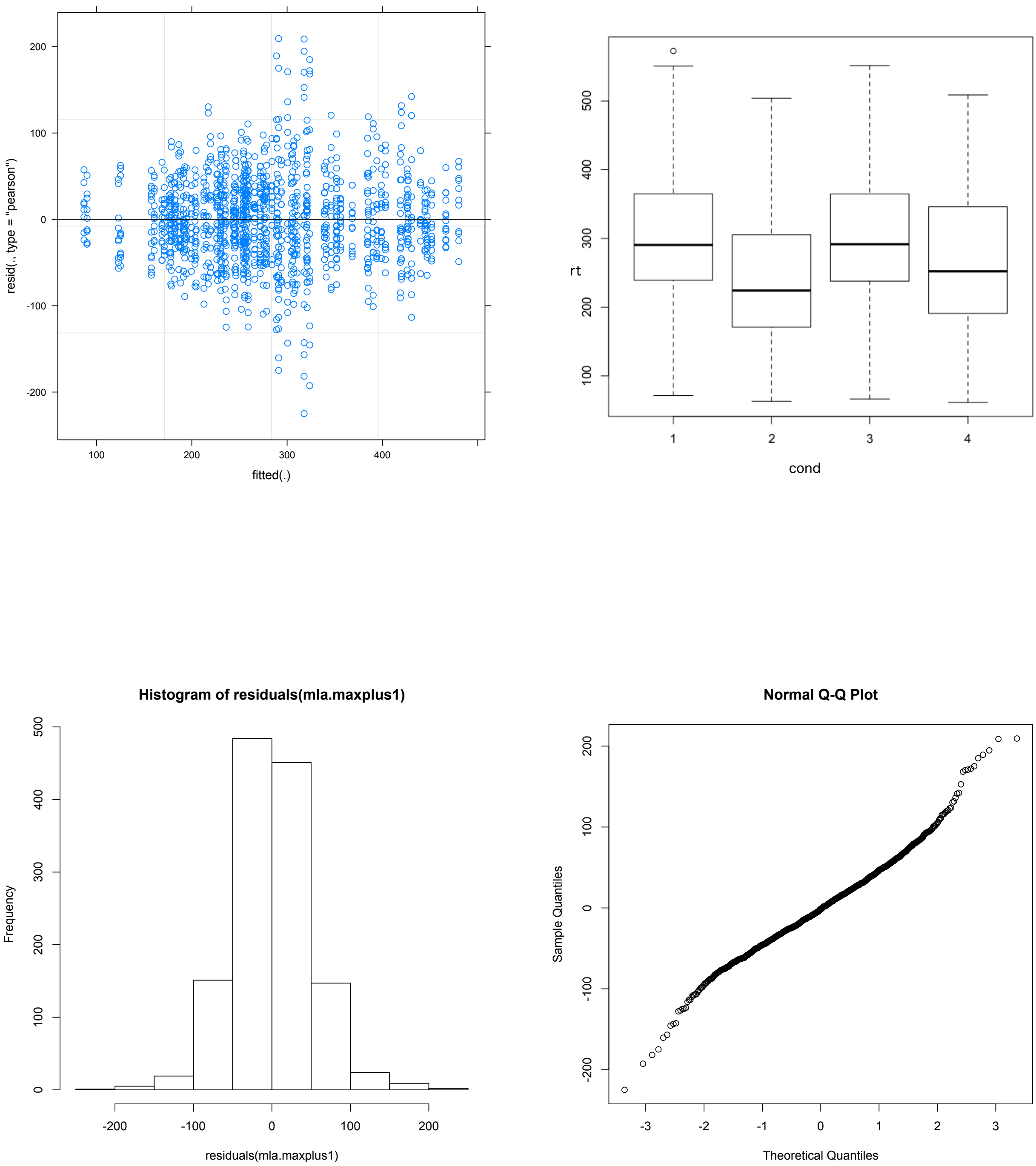


Figure 8.8: Residuals Analyses: Experiment 8

Table 8.7: Mean Naming Latencies (in ms): Experiment 8

	(1) Compounds	**(2) Phrases**	**(3) Disyll**	**(4) Mono**
PWds	1	2	1	1
LexWs	2	2	1	1
Syllables	2	2	2	2
Mean Lat. (ms)	301 (8.7%)	240 (7.7%)	300 (8.9%)	267 (6.3%)

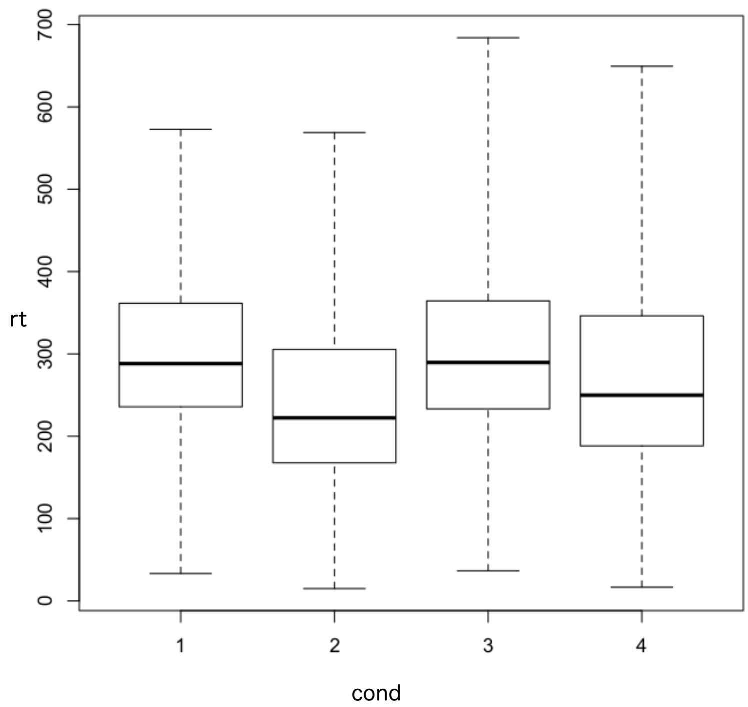


Figure 8.9: Mean RTs by Condition: Experiment 8

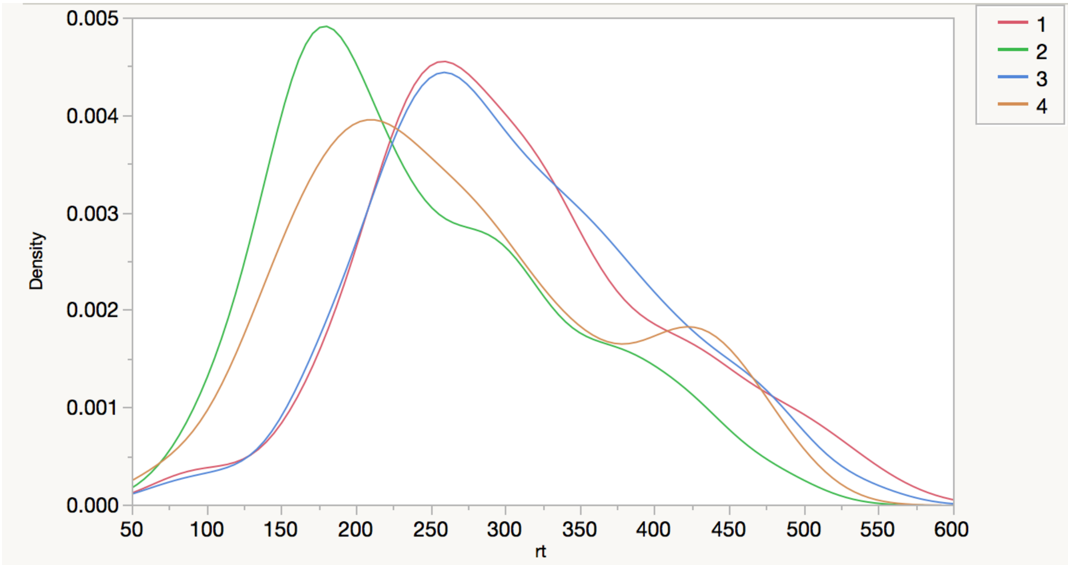


Figure 8.10: Densities of RTs by Condition: Experiment 8

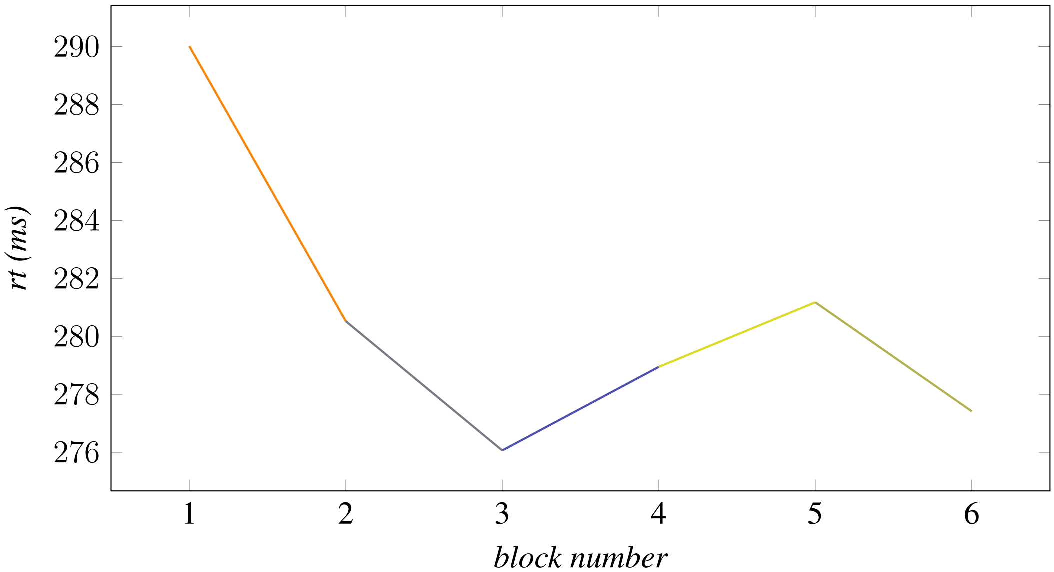


Figure 8.11: Mean RTs by Block: Experiment 8

8.3.3.2 Errors

As part of this investigation, an error analysis was carried out: errors were categorised as "time out" (the subject said nothing), "voice key error", "disfluencies" (e.g. stuttering), "stress error" (incorrect prosody), and "wrong item". The error rates were analysed proportionally using a generalised linear model (GLM) with a binomial distribution (link= logit).The analysis revealed no main effect of condition on error rate (χ^2= 5.44 p= 0.14). There were no other significant effects.

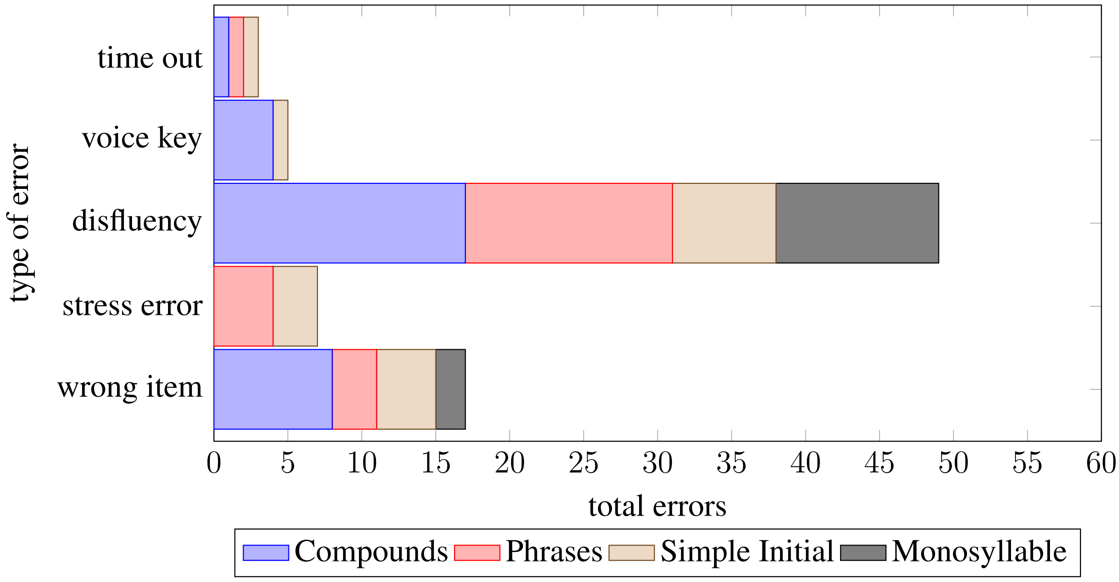


Figure 8.12: Total Errors: Experiment 8

8.3.3.3 Frequency and Word Length Analyses

For this analysis, the RT data were again submitted to a linear mixed models analysis, in which frequency and condition were all treated as the fixed effect factors. Once again, the maximally-appropriate structure that converged presented random intercepts and slopes for subject factors and a random intercept only for the by-items factor.[8] The models converged and their results are shown in Table 5.9 below. Neither the log (Estimate= 15.90, SE= 17.88, t= 0.889) nor CELEX (Estimate= 5.122, SE= 5.639, t= 0.908) measures of total word frequency had an effect on naming latencies in this task. In the compound frequency analysis, neither the first (Estimate= 0.01206, SE= 0.02027, t= 0.595) nor second (Estimate= 5.62403, SE= 1.01002, t= 0.555) morpheme word frequency affected naming latency, nor did the the first (Estimate = -0.016453, SE= 0.009209, t= -1.787) nor second (Estimate= 9.457e-03, SE= 1.00602, t= 0.94) morpheme

[8]Formula: rt $\sim$ cond + freq + (1 + cond | sub) + (1 | item).

of the adjective-noun phrase.[9] Lastly, an analysis of the effect of word length (number of letters) resulted in a significant value for words containing 7 and 8 characters (ts > 2.2), but no others. There was no interaction between length and condition.

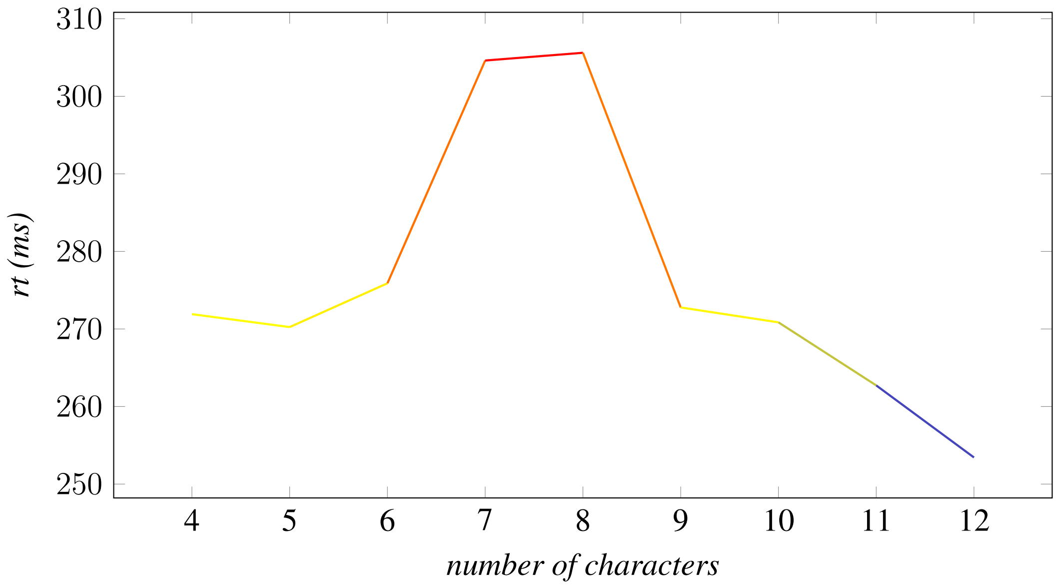


Figure 8.13: Mean RTs by Word Length: Experiment 8

8.3.4 Discussion

This experiment elicited a significant main effect of condition on naming latencies. Adjective-noun phrases, when paired with the auxiliary *are*, elicit significantly shorter naming latencies than either compounds, disyllabic words. or monosyllabic words. A pairwise comparison revealed significantly different naming latencies for these items, as well as the monosyllabic + clitic targets. Adjective-noun phrases resulted in the fastest responses, followed by monosyllabic + clitics, and compounds and disyllabic words elicited similar naming latencies to one another when paired with an auxiliary. This was in line with the results we saw from the native English speakers, suggesting that the naming latencies for both groups of speakers were affected by the size of the first prosodic unit:

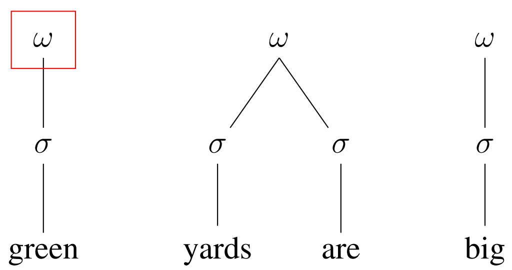
[9] because only 5 adjectives were used in these targets, the model was rank-deficient.

Table 8.8: Size of the Prosodic Unit in Experiment 4

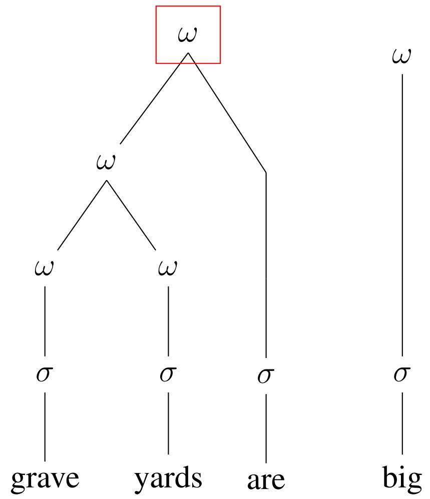
Cond	First PWd Unit	Size of First PWd Unit
Compound + Clitic	$(\text{doorways are})_\omega$	$(\sigma\ \sigma\ \sigma)$
Disyllabic + Clitic	$(\text{donkeys are})_\omega$	$(\sigma\ \sigma\ \sigma)$
Monosyllabic + Clitic	$(\text{dates are})_\omega$	$(\sigma\ \sigma)$
Adj-N + Clitic	$(\text{dark})_\omega$	(σ)

In the adjective-noun phrase condition, speakers only had to plan a monosyllabic adjective (e.g. *dark*). The monosyllable word + clitic condition consisted of two syllables when the auxiliary attached leftwards to it, eliciting slightly longer naming latencies than the adjective of the phrasal condition. Compounds and disyllabic words appeared no different in structure: the auxiliary attached leftwards to both targets the same.

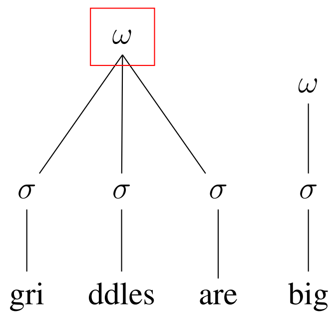
(8.6) Adjective-Noun Phrases + Clitics: Experiment 8



(8.7) Compounds + Clitics: Experiment 8



(8.8) Units predicting naming latencies in simple words: Experiment 4



These results indicate that the highly-fluent L2 speakers behaved similarly to the native English speakers in regards to the planning of the compound and phrasal conditions.

Returning to our predictions for this task, we find that the loss of planning time has elicited the same effect as in previous online experiments: speakers were only able to encode the first unit of the utterance before initiating speech. Crucially, this unit remains a well-formed Phonological Word, even when speakers are under additional pressure to answer each prompt with the correct construction in their L2 language. As in the other online tasks, reaction times were affected by experimental blocks: however, in this task, reaction times became **faster** over the course of the experiment. This was not isolated to this task: the non-native speakers also became faster in the middle blocks of Experiment 6 (the online compound naming task). Errors in this task were predominantly of the "disfluency" type, and the compound condition elicited the highest number of errors. However, errors were not particularly numerous for this task. As with the other tasks in this group, word frequency did not elicit any type of effect on reaction times in this task.

8.4 General Discussion

Taken together, these experiments have revealed some interesting details about how non-native speakers of English treat both auxiliaries and compounds. As mentioned in the introduction to this chapter, we know very little about the preparation of complex morphosyntactic words in L2 speakers. Having established the native English speakers treated compounds and single prosodic units (at least for the purpose of phonological planning), we sought to examine how highly-fluent L2 speakers might do the same. Results from Experiments 5 and 6 indicated that the Bengali speakers were able to cope quite well with the prosodic framing in English, but only if they were fluent enough. We also elicited evidence that certain features of the non-native language were activated at the phonological level: this was evident from the significant increase in errors related to the iambically-stressed condition in the delayed compound naming task. Following this, we hypothesised that Bengali speakers would again be able to access the correct prosodic frames for both compounds and noun phrases in English under more complex task conditions. However, we were unsure about how they might treat English clitics.

As discussed in §8.2.1, Bengali speakers are very familiar with the phenomenon of cliticisation, particularly in regards to compounds. Experimental evidence for the treatment of auxiliaries in non-native speakers of English is sparse: many studies that discuss this phenomenon are heavily weighted towards language-learning, and therefore suffer from a lack of sufficient theoretical background. Gut (2003) found that L2 speakers tended to reduce auxiliary and function words less than L1 speakers. Perales and Liceras (2010) also found that young L2 speakers failed to incorrectly inflect the auxiliary *do* to *don't* in negative clauses. Following this, we first questioned whether the Bengali speakers would reduce the English auxiliary *are* at all.

The results were surprising: in both experiments, it appeared that the Bengali speakers had little trouble producing regular auxiliary deletion. This was confirmed in the online task results, where the size of the first Phonological Word unit affected the naming latencies. Monosyllabic word targets, when paired with clitics, took significantly longer to prepare than monosyllabic words alone (Condition 4). Not only were these speakers able to access the correct prosodic framing for the target word + clitic condition when they had time to plan, they were also able to do so when planning time was removed.

As expected, both experiments resulted in longer naming latencies for the L2 speakers. This is in line with previous findings from dual-language interference tasks, e.g. Costa et al, 1999. Predictably, the size of the main effect of condition type was smaller for the Bengali speakers than it was for the native English speakers in the delayed experiment: they exhibited a difference in naming latencies of compounds and phrases of around 20 ms, while the native English speakers showed about 70 ms between the two. However in the online task, the effect was actually larger for the L2 speakers: compounds only differed from phrases in native speakers by 25 ms, whereas L2 speakers elicited an average of a 40 ms difference. This is also contrary to the results of Experiment 6, where both speaker groups generated roughly the same effect size (30 ms in L2 speakers, 39 ms in L1 speakers). This leads us to consider task type once again. We discussed the differences between target utterances between the compound tasks and clitic tasks earlier in this study; we theorised that the multi-word utterances of the clitic experiments may actually be closer to the process of generating natural language than the compound experiments, particularly for the online tasks. These results certainly support such an argument: rather than completing a simple reading task, the speakers had sufficiently more to do in the online clitic task. Therefore, effect size for the L2 online clitic conditions reflects the difficulty of the task.

To conclude, this chapter has provided us with more evidence for the planning unit in L2 English speakers: when possible, they endeavour to plan compounds as single recursive Phonological Words that contain clitics. L2 speakers, despite their previous troubles in assigning stress had no trouble allowing the auxiliary *have* to reduce and collapse onto the neighbouring Phonological Word. Crucially, in most cases, it also attached leftwards. Again we step back to take a look at the group of results together. There is overwhelming evidence here of the difficulties of speaking in the non-native language: in Experiment 5a and 5b, we saw the effect that fluency had on the entire task. Highly-fluent speakers were able to identify compound words and, most importantly, treat them differently than phrases for the purposes of phonological encoding. Conversely, the low-fluency group produced statistically-similar naming latencies across all four conditions, suggesting that, at the very least, they could not differentiate between compounds and phrases. Error analyses revealed some very interesting results regarding planning: unlike the L1 speakers, who mostly made errors of disfluency, the non-native speakers regularly made 'wrong item' or 'time out' errors. Returning to our second research question, we find evidence in all four tasks that the planning unit, at phonological encoding stage, retains a prosodic shape in non-native speakers of English. As the native speakers sacrifice initiation speed to prepare well-formed Phonological Words in the online tasks, so must the non-native speakers do so in both delayed and online tasks. Crucially however, in all cases, the unit being planned is the recursive Phonological Word.

9

Conclusion

9.1 Introduction

These eight experiments sought to answer the question, "what is the unit being planned during the prosodification process in phonological encoding?" in the English spoken by native and non-native speakers. We aimed to isolate more of the post-lexical processes involved in language production. Models of language production have shared the same three stages since their inception- even the early localisationist accounts presented a multi-stage process dedicated to the generation of ideas, words, and speech (e.g. Kussmaul, 1887; Lichtheim, 1885). Out of the nineteenth-century speech impairment studies grew a rich tradition of modelling language: we saw how speech errors formed a significant foundation of our early understanding of how speech is formed. As evidenced by the error analyses in this study, speech errors continue to be an important source of information about certain processes involved in speech production: in particular, we saw that the errors made by non-native English speakers differed in very specific ways from the native speakers (e.g. more lexical errors and more time outs). Recent language production models are remarkable in that no longer view the post-lexical encoding stage as a single process; rather it consists of several sub-processes all devoted to preparing a specific part of the post-lexical representation. Furthermore, they no longer assume a lexically-shaped unit as the underlying representation during phonological encoding. In this chapter, we summarise our findings for this study and discuss, in detail, their implications for both native and non-native post-lexical speech processing.

9.2 Summary of Findings

9.2.1 Native Speaker Results

We began this investigation by establishing two goals. The first goal was to elicit experimental evidence for the planning of connected speech at the post-lexical encoding stage. Both Levelt (1989, et al. 1999) and Roelofs (1997) have agreed that the unit being prepared at this stage is the Phonological Word. Evidence from experimental studies on production in Dutch and European Portuguese also supports this claim. Following this, we hypothesised that English speakers would also plan Phonological Words at this stage in phonological encoding, however little evidence exists for this process in English. Therefore our first task was to test this hypothesis: we ran a delayed speech production task based on Sternberg et al. (1978) and modelled after Wheeldon & Lahiri (2002) to test prosodification in English. We compared naming latencies for noun-noun compounds, adjective-noun phrases, and monomorphemic nouns and found that naming latencies for the compounds were significantly shorter than the phrases, and similar to the latencies for the monomorphemic words. This suggested that English speakers behave the same as those in the other studies: but evidence from online planning experiments (cf. Wheeldon & Lahiri, 1997) indicate that the shape of the planning unit might change under different task conditions. Consequently, we ran an online version of the production task in Experiment 1: here we found that English speakers still planned prosodic units during online planning. However, they only planned the first Phonological Word unit, and the naming latencies in this task depends on the **size** of this unit.

The first set of experiments established that the planning unit in English speakers was also a Phonological Word. Crucially, we found that speakers treated compound words as single prosodic items, at least for the purposes of planning spoken utterances. In order to pursue this hypothesis further, we tested how these units held up in multiword utterances, and crucially, with cliticisation. Experiments 3 and 4 tested naming latencies for utterances containing the original four conditions, with the added feature of an auxiliary, e.g. *dishcloths are clean.* We hypothesised that, if English compounds are really acting as single, recursive prosodic units, then the auxiliary should reduce and attach to the entire unit in normal speech, e.g. *dishcloths're nice.* Once again the task was run under both planning conditions: delayed and online. In the delayed task, we found that speakers treated noun-noun compounds similar to monomorphemic words, while latencies were significantly longer for the phrasal conditions. This suggested that the auxiliary was attaching to the entire prosodic unit of the compound, yet only attaching to the noun of the adjective-noun phrase in the phrasal condition. In the online speech task, results again suggested that planning was restricted to the size of first phonological unit: in this case, naming latencies for noun-noun compounds were similar to disyllabic monomorphemic words, while adjective-noun phrases generated the shortest naming latencies (due to the fact that speakers were only accessing the first unit of the phrase- the monosyllabic adjective).

The results of both groups of experiments revealed some important information about phonological encoding- and prosodification, in particular- in native English speakers. We confirmed that speakers were indeed generating prosodic frames according to Phonological, not lexical, Words. We were also acutely aware of the issue of frequency in compound planning and production. On account of this, frequency was controlled for, and the lack of frequency effect in the native speakers' results showed that word frequency was not affecting naming latencies. The difference in naming latencies between compounds and phrases indicated that the prosodic frames for the complex morphosyntactic structures of compounds were constructed according to the outer, recursive Phonological Word structure of the compound. This was remarkable as it revealed both experimental evidence for the planning of prosodic structures in English, and evidence for the makeup of the structure itself.

9.2.2 Non-Native Speaker Results

The second goal was to examine this process in non-native speakers of English. Once we had established that native English speakers were planning Phonological Words, we turned to a much less-known area of language production: the preparation of speech in a non-native language. Early attempts to model non-native language production focused primarily on the representation of lexical forms and the structure of the mental lexicon. Experimental data has indicated that learning a new language requires the neural systems related to linguistic processing to adapt in order to cope with a new syntax and lexicon. We also know that age, method of instruction, and modality (i.e. whether the language is signed or spoken) are all factors in how proficient a speaker becomes in their L2. However very little is known about the preparation of phonology in non-native speakers: the majority of studies of L2 planning have been firmly focused on acquisition and proficiency. A notable and welcome departure from this lies in Roelofs (2003; & Verhoefs, 2006), who found evidence that the phonological representations of both the L2 (target) and L1 languages were being activated during L2 production: that is, speakers exhibited shorter naming latencies when planning a word that had common initial segments in both languages. This, they argued, suggested that phoneme representations were shared amongst languages. Taking this into consideration, Roelofs & Verhoefs (2006) found that, with a few adjustments, the WEAVER++ model was able to adapt to non-native speech production without much trouble. One of the conditions of WEAVER++ is that only the phonological representations linked to a particular lemma are activated during word-form encoding. In Roelofs & Verhoefs (2006), they loosen this requirement, they allowed the cascading activation of non-activated lemmas from lexicalisation level to word-form level. In addition to this, the non-native WEAVER++ model introduces control processes: when speaking in an L2, the speaker places a goal symbol in working memory which specifies to speak in the L2); this goal symbol is activated at every stage of production.

In Experiments 1 and 2, we established that native English speakers planned nominal compounds as single prosodic units. Experiments 3 and 4 provided additional evidence for this, in that clitics attached to compounds in the same manner as monomorphemic words. Using the same tasks, we tested the planning of compounds in L2 English speakers. We used native speakers of Kolkata Bengali as our non-native speaker group, both on account of their language background and the shared features of the two languages. Compounds composed of two or more noun stems are common in Bengali, and cliticisation occurs regularly throughout the language. Our initial questions for the non-native speaker groups were slightly different to those for the English speakers. First, we wanted to know if the prosodic frames were constructed the same way: would non-native speakers recognise compound words as single prosodic units? Or would they treat them as two lexical words? And, more importantly, would fluency in English affect this process?

Following this, Experiment 5 tested delayed planning in two groups of non-native English speakers: highly-fluent and less-fluent. We found that the highly-fluent L2 speakers generated results similar to L1 speakers: naming latencies for compounds were identical to the monomorphemic word conditions. Phrases again generated much longer response times. The less-fluent L2 speakers, however, generated flat results across all four conditions: whether or not this was due to recognition failure, retrieval failure, or encoding failure was unclear. Crucially, however, naming latencies for the compounds and phrases were identical for these speakers. Furthermore, the number of errors were in the hundreds.

In regards to errors, the highly-fluent speakers also generated a high number of errors; this was due, however, not to a lack of fluency, but to a problem associated with one of the control conditions. This task used disyllabic words as control conditions: one condition contained words with initial stress, and the other with final stress. Bengali does not have final stress: as Professor Chatterji noted in 1926, stress is overwhelmingly assigned to initial syllables both in words and phrases. It is little wonder, therefore, that this condition resulted in over 300 errors (out of 900 trials). We theorised that a combination of the lack of word recognition and default Bengali stress patterns resulted in these errors; simply put, the speakers may not have been familiar with the English word enough to know to activate the iambic stress associated with it. This resulted in errors such as *gázèlle* for *gàzélle* and *lápel* for *làpél*. Once this condition was removed (as it was in Experiment 6), the number of errors generated by the highly-fluent speakers fell significantly. Furthermore, the results in the online task were similar to those of the L1 speakers: highly-fluent L2 speakers also sacrificed utterance initiation speed in order to plan well-formed Phonological Words. Crucially, they again treated compounds as single prosodic units.

As in the native speakers, the next set of experiments tested the planning of compounds and clitics in L2 speakers. There was little indication from Experiments 5 and 6 about which way the auxiliaries might attach- if they even did attach. Our questions for these experiments were thus: how do L2 speakers treat auxiliaries in English? Previous research has indicated that non-native speakers are less inclined to exhibit reduction and deletion of vowels in spoken language (cf. Gut, 2003). Delayed task results for our speakers initially suggested that the auxiliary "are" was in fact exhibiting reduction and attachment; however, because the delayed task indicated the total number of Phonological Words being prepared by the speakers, it was still unclear exactly how the auxiliary was being treated. The real test came in the form of the online task: because speakers were only able to plan the first prosodic unit of the utterance, this would indicate the size of the unit. If the speakers were truly reducing the auxiliary, then this effect would be translated into the naming latencies for each condition. Accordingly, results for the online task were supportive of both reduction and leftwards attachment in the Bengali speakers. The mean naming latency for noun-adjective phrases was 242 ms, with monosyllabic simple words at 286 ms, and compounds and disyllabic simple words showing similar results at 306 ms. This suggested to us that speakers were accessing the English prosodic framing, and that this framing treated compounds as single prosodic units.

Table 9.1: Summary of Findings

Experiment	Stimuli	Task Type	Speaker Group	Mean Naming Latencies and Error Percentages
1	compounds	delayed	English	Compounds: 445 ms (6.6%) Phrases: 485 ms (4.2%) Disyllabic Initial Stress: 442 ms (6.6%) Disyllabic Final Stress: 449 ms (6.2%)
2	compounds	online	English	Compounds: 262 ms (8.3%) Phrases: 223 ms (8.6 %) Disyllabic Simple: 259 ms (7.0%) Monosyllabic Simple: 221 ms (14.3%)
3	compounds + clitics	delayed	English	Compounds: 331 ms (6.3%) Phrases: 400 ms (4.2%) Disyllabic Simple: 326 ms (5.1%) Monosyllabic Simple: 332 ms (5.6%)
4	compounds + clitics	online	English	Compounds: 215 ms (4.2%) Phrases: 190 ms (3.2%) Disyllabic Simple: 222 ms (4.0%) Monosyllabic Simple: 208 ms (4.3%)
5a	compounds	delayed	Bengali (HF)	Compounds: 455 (10.1%) Phrases: 471 ms (13.1%) Disyllabic Initial Stress: 452 ms (17.0%) Disyllabic Final Stress: 450 ms (36.9%)
5b	compounds	delayed	Bengali (LF)	Compounds: 562 ms (18.0%) Phrases: 582 ms (16.3%) Disyllabic Initial Stress: 573 ms (19.3%) Disyllabic Final Stress: 552 ms (47.2%)
6	compounds	online	Bengali (HF)	Compounds: 204 ms (9.8%) Phrases: 174 ms (7.3%) Disyllabic Simple: 193 ms (9.2%) Monsyllabic Simple: 174 ms (6.0%)
7	compounds + clitics	delayed	Bengali (HF)	416 ms (8.5%) Phrases: 438 ms (5.0%) Disyllabic Simple: 414 ms (6.2%) Monosyllabic Simple: 419 ms (8.1%)
8	compounds + clitics	online	Bengali (HF)	Compounds: 301 ms (8.7%) Phrases: 240 ms (7.7%) Disyllabic Simple: 300 ms (8.9%) Monosyllabic Simple: 267 ms (6.3%)

9.3 Planning and Morphosyntactically-Complex Words

9.3.1 The Post-Lexical Shape of Complex Words

The survey of language production models in Chapter 3 focused on the evolution of the post-lexical processes, and in particular, the generation of prosodic frames. In short, we wanted to isolate the processes that prepare the features of spoken language. However, in order to fully understand what is happening at this level, we found that it was necessary to consider how items enter the post-lexical encoding stage.

In each experiment's discussion section, we made continuous references to one speech production model in particular: WEAVER++. WEAVER++ assumes a spreading-activation process for lexicalisation: lexical concepts activate lemma nodes, which in turn activate the item's syntax and grammatical environments. For example, English verb lemmas activate information which corresponds to the syntactic information of the word: number, person, tense and mood (Levelt et al., 1999: 4). This gives rise to certain questions regarding the shape of morphosyntactically-complex words in both the lexical and post-lexical encoding stages. How are compounds stored in the mental lexicon? Do they enter the post-lexical encoding stage as a single lexical unit?

We briefly discussed theories of compound lexicalisation in Chapter 4 and saw that experimental evidence for the storage and access of compounds at this stage is, for lack of a better term, inconsistent. Psycholinguistic studies designed to test compound frequency as evidence for the lexical status of these items often return contradictory results: some (e.g. Bien et al., 2005) found that only the summed frequency of a compound word had an effect on reaction times, while others (Juhasz et al., 2003; Duñabeitia et al., 2007) only found an effect for the final constituent of the compound word. Levelt et al. (1999) acknowledge the issues associated with complex morphosyntactic structures:

> "The case is more complicated for complex derivational morphology. Most of the frequently used compounds are of the type discussed here. For example, *blackboard, sunshine, hotdog*, and *offset* are most likely single lemma items,

though *thirty-nine* and complex numbers in general... might not be."

They envision the lexicalisation of morphologically-complex words as consisting of different routes: therefore, newly-formed compounds are formed by different routes than frequently-used compounds. Therefore, "lexicalised" compound words such as *blackboard* contain a single lemma node, but multiple form nodes for each morpheme in the compound. Newly-formed (or unknown) compounds, however, retain their individual lexical concepts and lemma nodes in WEAVER++. If this is the case, and familiar (high-frequency) compounds enter the phonological encoding stage already specified as single lemma nodes, is their internal prosodic structure actually any different than morphologically-simple disyllabic words?

Crucially, WEAVER++ conducts syllabification across word and morpheme boundaries, and morphemes are considered units of planning when they determine the shape of the word forms, e.g. their syllabification. Levelt et al. maintain that this is independent of semantic transparency in compound processing (1999: 28). Therefore, both opaque and transparent compounds undergo the same process at this post-lexical encoding stage. Following this, when a newly-formed compound exits the lexicalisation stage, it is still composed of separate lemma nodes. Roelofs et al. (1998) obtained evidence of morphological priming, and the size of the effect was identical for both opaque and transparent compounds. This suggests that the default prosodic frame for both types of compounds requires the internal boundaries to be present in the post-lexical encoding stage (Levelt et al., 1999: 28).

9.3.2 Prosodic Frames and Post-Lexical Encoding

We hypothesise that prosodic frames are built according to a unit defined in prosodic constituent theory as the Phonological Word: evidence for this unit comes both from a wealth of theoretical and experimental studies. We saw glimpses of its inception in the nineteenth-century grammars of Sweet, Saran and Sievers, in which the scholars grappled with the differences between the lexical (or syntactic) structure of language and its spoken organisation. Franz Saran, in particular, theorised that language could be organised into hierarchically-ordered groups based on their sound features (e.g. loudness, strength, and syllabification).

The generative tradition gave prosodic constituent theory its modern foothold. In *The Sound Patterns of English*, Chomsky and Halle argued that the components responsible for generating syntax and phonology were separate and, like the early grammarians, part of their argument hinged upon the fact that the syntax and phonology of language did not always match up. Word stress in English, in particular, received a great deal of attention in regards to this claim, first by Liberman and Prince, then Selkirk in 1979 and 1980. It was within Liberman and Prince's 1976 publication that the first tangible theoretical framework of prosodic constituent theory was built. They argued that stress assignment in English was more complex than Chomsky and Halle initially proposed: for one, multi-syllabic words such as *execution* and *revolutionary* required more than just "strong" and "weak" stresses. Syllables were not simply stressed, or unstressed: when uttered in regular speech, there are varying degrees of intensity on the syllables. Furthermore, they noted that compounds and phrases that shared the same words (e.g. *blackbird* and *black bird*) differed only in stress assignment; there needed to be a way to depict this at a level higher than the individual syllables. In their theory, stress was assigned to nodes at different levels in the tree. These levels included syllable nodes, metrical foot nodes, and even a level dedicated to multi-word structures (such as *labor union*: the "mot" level. Selkirk took Liberman and Prince's levels and translated them into individual hierarchically-organised structures called constituents. Each constituent was subject to specific rules and processes associated with its domain, and the constituent responsible for some of the most notable features of spoken language is the Phonological Word constituent.

A Phonological Word can constitute a single stressed foot (e.g. *cat*) or a lexical word combined with any number of unstressed function words. Cliticisation, affixation, and auxiliary reduction often results in this second definition, particularly in spoken English where auxiliaries often lose stressed status and collapse onto an adjacent word (e.g. *what are* > *what're*). In Chapter 2, we discussed at length the specific controversies related with both cliticisation and compounding. We saw that, in the end, these structures require a recursive treatment: there was significant experimental and empirical evidence that compound words, in particular, are treated like single units in spoken language. Although it is a valid and long-established operation at the morphosyntactic level, recursion has been stoutly rejected from operating at the phonological level (cf. Jackendoff and Pinker, 2005). However, there have been arguments for recursion in phonology, particularly above the word level (Ladd 1986, Hunyadai 2010, van der Hulst, 2010 and Vigario, 2010). And while phonology itself may reject recursion, the constituents may simply be marked for it: Selkirk argued that recursion at the phonological level simply mirrors the recursive structure of the morphology and syntax.

Analyses that outright reject recursion meet with problems in prosodic constituent theory: we saw Nespor and Vogel's various attempts to account for the structures of the controversial compound and clitic units. First they assigned clitics to a special constituent level (the Clitic Group) and when that failed, they assigned both compounds and clitics to another special level (the Composite Group). Both new constituents were proposed in order to avoid breaking the rules of the Strict Layer Hypothesis, which stipulates that prosodic constituents cannot dominate themselves. But, as we saw in the "hog dog buns are tasty" tree structures, these constituents still involve either recursive domination or other rule violations (e.g. the clitic attaching at a higher level, an operation also rejected by the SLH). Instead of introducing new constituents to deal with these troublesome structures, Selkirk revisited the SLH by applying the theoretical framework of Optimality Theory (Prince & Smolensky, 2004). Instead of acting as a single, all-encompassing rule, the SLH was fleshed out into individual, violable constraints on layeredness, headedness, exhaustivity, and nonrecursivity. In line with the OT tradition, a constraint could be violated if it resulted in an optimal output representation. Therefore, the SLH constraint of nonrecursivity could be violated if a recursive structure was the best possible choice for dealing with a particular constituent (as in the case of the Phonological Word). Applying the SLH as a series of violable restraints does not weaken it; in fact, it only strengthens our ability to cope with the phonological behaviour of auxiliaries and compounds. Under the new constraint-based SLH framework, we allow for an outer (or higher) Phonological Word constituent to dominate lower ones: in the case of compounds, this dominates two (or more) individual Phonological Words that make up the compound. In the case of clitics, this dominates both the lexical word (which by default is a Phonological Word) and any unstressed items that may attach to it.

In the results of the experiments presented in this thesis, we have seen evidence that supports recursion at the Phonological Word level. In every instance, native English speakers treated compounds as single prosodic units. Naming latencies for these items were significantly-different to the phrases, even when the complexity of the utterance was increased. When used in the company of auxiliaries in regular speech, the auxiliaries reduced and attached to the compounds as a whole, not to the individual constituents of the compound words. This adds strength to an argument in favour of the existence of an outer, or higher, recursive Phonological Word, and more importantly, that this is the structure around which prosodic frames are built during the phonological encoding process of language production.

9.4 Planning and Non-Native Speech Processing

9.4.1 Post-Lexical Processing in Non-Native Speakers

In the hypotheses for the non-native tasks, we theorised that the results would exhibit certain "markers" of L2 naming, e.g. an increase in errors, an increase in overall latencies, and a decrease in effect size. In Experiment 5a (delayed naming of compounds), we saw a very large number of stress errors associated with the finally-stressed control condition (Condition 4): over 300 errors out of 900 total trials. This was due to the existence of the Bengali stress rule, which stipulates that stress in Bengali is overwhelmingly word-initial. WEAVER++ assumes that certain aspects of L2 phonological encoding cannot be shared: for example, languages differ in terms of rules of syllabification and stress assignment. Following this, we saw that, while the Bengali speakers were able to recognise the orthographic representation and activate the associated phonology of the English word, their familiarity with the target word did not extend to stress placement. In most cases, the subjects produced the correct word, only with an incorrect stress pattern (e.g. *gazelle* for *gazelle*). Therefore, the encoding process simply defaults to the stored prosodic shape for disyllabic words.

We also found that speakers tended to replace less familiar words with higher-frequency items, e.g. *bell* (CELEX 745) for *ballad* (CELEX 50). These findings fit with the predictions of the WEAVER++ model of L2 production: if the L2 word has less use (or has been learned later in life), it will likely receive less activation; if the L1 and L2 languages are unbalanced to begin with, all L2 lemmas are already receiving less activation. The activation effect is therefore magnified in non-native speakers, resulting in words with lower frequency generated significantly longer reaction times in our data. Furthermore, these items almost always shared initial phonological segments, lending support to the incremental activation and planning of segments at this stage. These results further cement the need to treat the post-lexical encoding process in language production as a series of individual subprocesses that plan their own representations. And as we will see below, an error during one subprocess does not necessarily mean the entire post-lexical encoding process will be unsuccessful in an L2 speaker.

9.4.2 Fluency and Post-Lexical Processing

One of the goals of this study was to isolate and examine prosodification in non-native speakers of English. Over the course of these experiments, we found that our non-native speakers were able to plan and encode complex prosodic structures without much difficulty as long as they were able to identify them. L2 speakers of English exhibited a flexible planning scope, as long as they were fluent enough to complete the task. While the planning scope varied depending on the task conditions (i.e. delayed or online), the unit planned by non-native speakers was the same as in native English speakers: the Phonological Word.

Nevertheless, we saw in Experiment 5b that fluency had a serious effect on these tasks: the less-fluent native speakers were not able to distinguish between any of the four conditions. Simply put, the tasks did not work if the subject did not have a suitable level of fluency in English. The term "fluency" is one whose definition is not easily obtained; most often, it is simply used to reflect oral proficiency and accuracy in the L2 language. However, as we saw in the highly-proficient results discussed above, the L2 speakers produced similar results to L1 speakers in regards to the treatment of compounds and phrases. Crucially, the overall mean naming latencies for the disyllabic final-stressed condition did not differ at all, despite the high number of errors. In the clitic + compound experiments, the L2 speakers performed in an entirely contradictory manner to expectation: they did not produce less function word reduction. In both experiments, it appeared that the Bengali speakers had little trouble producing regular auxiliary deletion. This was confirmed in the online task results, where the size of the first Phonological Word unit affected the naming latencies.

These results indicate that fluency, as we conceive it, reaches far beyond simple oral proficiency and accuracy. Fluency is also reflected in the ability to access the correct prosodic structure of the L1 language and the storage of non-default prosodic structures (i.e. with the final-stressed English words) within the lemma in the mental lexicon.

9.5 Conclusions and Future Directions

In both delayed priming tasks, we found overwhelming evidence that reaction times reflected the total number of prosodic units in the target sentence. In the online tasks, however, speech latencies only reflected the size of the first prosodic unit. Taken together, these results suggest that, despite containing two lexical and prosodic words, English compounds are planned as single prosodic units, exhibiting encliticisation and reaction times similar to those of monomorphemic words. As shown by the results in this study, this naming paradigm has proved extremely beneficial for eliciting data about the structure of prosodic units in speech production. Not only was it successful for native speakers of Dutch, European Portuguese, and English, we also found that it was easily implemented into a study of post-lexical encoding in non-native speakers of English.

Several results deserve further attention. The first is the disparity between the naming latencies for the online task in native speakers (Experiment 2) and those for L2 speakers in Experiment 6: here we found that naming latencies were shorter than those of native English speakers (compare 186 ms for this task to the 241 ms average for the native speakers). This is contradictory to the delay commonly associated with naming tasks in L2 speakers (cf. Gollan et al., 2005). This may be due to a number of factors, including 'over-rehearsing' from the L1 speakers or performance-related pressure in the L2 speakers. It is also worthwhile to note that this speed came at a cost; errors were significantly higher for the non-native speakers in the online task. Nevertheless, both groups of data deserve a closer look.

Secondly, as noted in Chapter 8, the Phonological Word in Bengali behaves much like its English counterpart, particularly in respect to clitic attachment. Bengali is very compound-productive, which may facilitate recognition of the prosodic structure of compounds in these speakers. Another next step would be to run these experiments on L2 speakers of English whose native language is not compound-productive, such as Hebrew or Norwegian. If the internal structure of compound words is not stored for these speakers, the results might differ significantly. On the whole, however, the findings of this study lend strong support to the claim that it is in fact the prosodic structure (not the lexical or morphosyntactic structure) of the utterance that is dictating the arrangement of prosodic frames during the post-lexical encoding stages of language production in English.

10

Appendix A

10.1 Language Background and Proficiency Tasks

Figure 10.1: The image originally presented here cannot be made freely available via ORA because of copyright. The image was sourced from Li et al., 2006.

11

Appendix B

11.1 Word Familiarity and Stress Questionnaires

Please read these instructions very carefully.

This survey is about word familiarity: how well you recognise and know the meaning of different words. Your task is to rate your familiarity with each word on a 5-point rating scale by selecting a number from 1 to 5. Use this guide to help you decide which number to select:

5- You recognise the word and are confident that you know the meaning of the word.

4- You think you know the meaning of the word, but are not certain that the meaning that you know is correct.

3- You recognise the word as one you have seen before, but you don't know the meaning.

2- You think you might have seen or heard the word before, but are not positive.

1- You have never seen or heard the word before.

We don't expect you to know or even recognise all of these words. Don't be concerned if there are many words you have never seen or heard before. These words reflect a very wide sample of words in English and it is expected that some of the words will be unknown to most people. Please be sure to respond to every word.

1. denim

1- not at all familiar	2	3	4	5- completely familiar
○	○	○	○	○

What syllable does the main stress (main emphasis) fall on?

(NATIVE BRITISH ENGLISH SPEAKERS ONLY)

1. daisy

○ first ○ second ○ not sure

2. brigade

○ first ○ second ○ not sure

3. dolphin

○ first ○ second ○ not sure

4. donkey

○ first ○ second ○ not sure

5. graphic

○ first ○ second ○ not sure

12

Appendix C

12.1 Proficiency Bands: IELTS

Figure 12.1: The image originally presented here cannot be made freely available via ORA because of copyright. The image was sourced from the IELTS Speaking band descriptors (public version), British Council.

References

Acheson, D., Ganushchak, L., Christoffels, I., and Hagoort, P. (2012). Conflict monitoring in speech production: Physiological evidence from bilingual picture naming. *Brain and Language*, 123(2):131–136.

Ackermann, H. and Riecker, A. (2010). The contribution(s) of the insula to speech production: a review of the clinical and functional imaging literature. *Brain Structure and Function*, 214(5-6):419–433.

Ahmadian, M. and Hosseini, S. (2012). A study of the relationship between Iranian EFL learners' multiple intelligences and their performance on writing. *Mediterranean Journal of Social Sciences*, 3(1):111–126.

Alario, F., Goslin, J., Michel, V., and Laganaro, M. (2010). The functional origin of the foreign accent: Evidence from the syllable-frequency effect in bilingual speakers. *Psychological Science*, 21(1):15–20.

Albert, M., Obler, L., and Press, L. (1978). *The bilingual brain: Neuropsychological and neurolinguistic aspects of bilingualism*. New York, London.

Anderson-Hsieh, J., Johnson, R., and Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Language learning*, 42(4):529–555.

Anderson-Hsieh, J. and Venkatagiri, H. (1994). Syllable duration and pausing in the speech of chinese esl speakers. *TESOL Quarterly*, 28(4):807–812.

Ando, E., Jared, D., Nakayama, M., and Hino, Y. (2014). Cross-script phonological priming with Japanese Kanji primes and English targets. *Journal of Cognitive Psychology*.

Aoyama, K. and Guion, S. (2007). Prosody in second language acquisition. *Language experience in second language speech learning: In honor of James Emil Flege*, 17:281.

Baars, B., Motley, M., and MacKay, D. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14(4):382–391.

Baayen, R., Davidson, D., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Baayen, R., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: evidence for a parallel dual-route model. *Journal of Memory and Language*, 37(1):94–117.

Baddeley, A. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *The Quarterly Journal of Experimental Psychology*, 18(4):302–309.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4(10):829–839.

Baddeley, A., Logie, R., Bressi, S., Sala, S., and Spinnler, H. (1986). Dementia and working memory. *The Quarterly Journal of Experimental Psychology*, 38(4):603–618.

Badecker, W. (2001). Lexical composition and the production of compounds: Evidence from errors in naming. *Language and Cognitive Processes*, 16(4):337–366.

Baginsky, A. (1871). Aphasie in Folge schwerer Nierenerkrankungen. *Berliner Klinische Wochenschrift*, 8:439–443.

Balota, D., Boland, J., and Shields, L. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, 28(1):14–36.

Balota, D. and Chumbley, J. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, 24(1):89–106.

Baroni, M. (2001). The representation of prefixed forms in the Italian lexicon: Evidence from the distribution of intervocalic [s] and [z] in northern Italian. In *Yearbook of Morphology 1999*, pages 121–152. Springer.

Bates, D. (2006). lmer, p-values and all that. [https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html].

Bates, D., Mächler, M., Bolker, B., and Walker, S. (in press). Fitting linear mixed-effects models using lme4.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Eigen, C., and Rcpp, L. (2015). Package 'lme4'. *convergence*, 12:1.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, 10(2):344–380.

Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology*, 3(01):255–309.

Benczes, R. (2004). On the analysability of English exocentric compounds. *Jezikoslovlje*, 5(1+ 2):1–21.

Berman, R. and Clark, E. (1989). Learning to use compounds for contrast: Data from Hebrew. *First Language*, 9(27):247–270.

Bernstein, P., Scheffers, M., and Coles, M. (1995). "where did i go wrong?" a psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1312.

Bialystok, E., Craik, F., Green, D., and Gollan, T. (2009). Bilingual minds. *Psychological Science in the Public Interest*, 10(3):89–129.

Bien, H., Levelt, W., and Baayen, R. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49):17876–17881.

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, 3(4):395–421.

Bock, K. and Griffin, Z. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129(2):177.

Boersma, P. and Weenink, D. (2013). Praat software. *Amsterdam: University of Amsterdam*.

Bonin, P., Peereman, R., and Fayol, M. (2001). Do phonological codes constrain the selection of orthographic codes in written picture naming? *Journal of Memory and Language*, 45(4):688–720.

Booij, G. (1984). Neutral vowels and the autosegmental analysis of Hungarian vowel harmony. *Linguistics*, 22(5):629–642.

Booij, G. (2005). Compounding and derivation: Evidence for construction morphology. In *Morphology and its demarcations: Selected papers from the 11th Morphology meeting, Vienna, February 2004*, volume 264, page 109. John Benjamins Publishing.

Booij, G. and Rubach, J. (2008). Morphological and prosodic domains in Lexical Phonology. *Phonology*, 1:1.

Box, G. and Draper, N. (1987). *Empirical model-building and response surfaces*. Wiley: New York.

Bresnan, J. and Kanerva, J. (1989). Locative inversion in Chicheŵa: a case study of factorization in grammar. *Linguistic Inquiry*, 20(1):1–50.

Brown, R. and McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5(4):325–337.

Brysbaert, M., Van Wijnendaele, I., and De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, 104(2):215–226.

Buck-Gengler, C., Menn, L., and Healy, A. (2004). What "mice trap" tells us about the mental lexicon. *Brain and Language*, 90:453–464.

Butterworth, B. (1983). *Language production*. Academic Press London.

Butterworth, B. (1989). Lexical access in speech production. In *Lexical representation and process*. The MIT Press.

Cattell, J. (1885). The inertia of the eye and brain. *Brain*, 8(3):295–312.

Chakraborty, R. (2011). Influence of L2 proficiency on speech movement variability: Production of prosodic contrasts by Bengali–English speakers. *Bilingualism: Language and Cognition*, 14(04):489–505.

Chatterji, S. (1926). *The Origin and Development of the Bengali Literature, Volume II*. Calcutta: D. Mehra, Rupa & Co (1975 edition).

Cholin, J., Dell, G., and Levelt, W. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1):109–122.

Cholin, J., Schiller, N., and Levelt, W. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, 50(1):47–61.

Chomsky, N. and Halle, M. (1968). The sound pattern of English. In *Studies in Language*. New York: Harper & Row.

Christoffels, I., De Groot, A., and Kroll, J. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, 54(3):324–345.

Cinque, G. (1993). A null theory of phrase and compound stress. *Linguistic Inquiry*, 24(2):239–297.

Cole, J. F. (1990). The minimal word in Bengali. In *The Proceedings of the Ninth West Coast Conference on Formal Linguistics*, pages 157–170.

Collins, A. and Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407.

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language*, 45(4):721–736.

Colomé, À. and Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1):96–109.

Condoravdi, C. (1990). Sandhi rules of Greek and prosodic theory. In Inkelas, S. and Zec., D., editors, *The phonology-syntax connection*, pages 63–84. University of Chicago Press.

Costa, A., Heij, W., and Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition*, 9(02):137–151.

Costa, A., Miozzo, M., and Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, 41(3):365–397.

Cutler, A. (1979). Contemporary reactions to Rudolf Meringer's speech error research. *Historiographia Linguistica*, 6(1):57–76.

Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we able to run any psycholinguistic experiments at all in 1990? *Cognition*, 10(1):65–70.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4):385–400.

Damian, M., Bowers, J., Stadthagen-Gonzalez, H., and Spalek, K. (2010). Does word length affect speech onset latencies when producing single words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(4):892.

Damian, M. and Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, 57(2):195–209.

Damian, M., Vigliocco, G., and Levelt, W. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, 81(3):B77–86.

Dash, N. (2015). *A Descriptive Study of Bengali Words*. Cambridge University Press.

Davies, R., Rodríguez-Ferreiro, J., Suárez, P., and Cuetos, F. (2013). Lexical and sub-lexical effects on accuracy, reaction time and response duration: impaired and typical word and pseudoword reading in a transparent orthography. *Reading and Writing*, 26(5):721–738.

Davis, C. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37(1):65–70.

De Bot, K. (1992). A Bilingual Production Model: Levelt's 'Speaking'Model Adapted. *Applied Linguistics*, 13(1):1–24.

De Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism*, 1(1):17–32.

De Bot, K. and Schreuder, R. (1993). Word production and the bilingual lexicon. *Studies in Bilingualism*, 6:191–214.

De Groot, A. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5):1001.

De Groot, A. and Nas, G. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30(1):90–123.

De Jong, N., Feldman, L., Schreuder, R., Pastizzo, M., and Baayen, R. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language*, 81(1):555–567.

Delazer, M. and Semenza, C. (1998). The processing of compound words: A study in aphasia. *Brain and Language*, 61(1):54–62.

Dell, G. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science*, 9(1):3–23.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3):283.

Dell, G., Juliano, C., and Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, 17(2):149–195.

Dell, G. S., Burger, L. K., and Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological review*, 104(1):123.

Dickerson, W. (1989). *Stress in the speech stream: The rhythm of spoken English*, volume 2. Univ of Illinois Press.

Dijkstra, T., Timmermans, M., and Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, 42(4):445–464.

Dijkstra, T. and Van Heuven, W. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, 5(03):175–197.

Dimitropoulou, M., Duñabeitia, J., and Carreiras, M. (2011). Two words, one meaning: Evidence of automatic co-activation of translation equivalents. *Frontiers in Psychology*, 2.

Drachman, G. and Malikouti-Drachman, A. (1994). Stress and Greek compounding. *Phonologica*, 1992:55–64.

Duñabeitia, J., Perea, M., and Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? evidence for morpho-orthographic decomposition. *Cognition*, 105(3):691–703.

Duyck, W., Vanderelst, D., Desmet, T., and Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, 15(4):850–855.

Egilsson, S. (1825-37). Fms. i-x. In Rask, R. and C.C. Rafn, e., editors, *Fornmanna sögur útgefnar eptir gömlum handritum að tilhlutun hins Norræna fomfræða félags*. Kaupmannahöfn.

Egilsson, S. (1851). *Tvö brot af Haustlaung og Þórsdrápa (Se. bls. sq. 61-64), farð til rètts máls, og útskýrð með glósum í stafrófsröð af Svb. Egilsson*.

Ellis, A. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior*, 19(5):624–634.

Faarlund, J. (2004). *The Syntax of Old Norse: With a survey of the inflectional morphology and a complete bibliography: With a survey of the inflectional morphology and a complete bibliography*. OUP Oxford.

Fabb, N. (1998). Compounding. In Zwicky, A. and Spencer, A., editors, *The Handbook of Morphology*. Oxford: Blackwell.

Ferrand, L. and Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the psychonomic society*, 31(2):119–122.

Ferrand, L., Segui, J., and Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, 35(5):708–723.

Ferrand, L., Segui, J., and Humphreys, G. (1997). The syllable's role in word naming. *Memory & Cognition*, 25(4):458–470.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30(2):210–233.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100(2):233.

Ferreira, F. and Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46(1):57–84.

Ferreira, V. S. and Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(6):1187.

Finkbeiner, M., Gollan, T. H., and Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and Cognition*, 9(02):153.

Fiorentino, R. and Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, 22(7):953–1000.

Foroodi-Nejad, F. and Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in persian–english bilinguals. *Bilingualism: Language and Cognition*, 12(04):411–427.

Fox, A. (2000). *Prosodic features and prosodic structure: The phonology of suprasegmentals*. Oxford University Press.

Fox, J., Bates, D., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Monette, G., Nilsson, H., Ogle, D., et al. (2009). "Car": Companion to applied regression, R Package version 1.2-16.

Francis, W. (1999). Cognitive integration of language and memory in bilinguals: semantic representation. *Psychological Bulletin*, 125(2):193.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, pages 27–52.

Fromkin, V. (1973). *Speech errors as linguistic evidence*. Walter de Gruyter.

Fromkin, V. (1977). Putting the emPHAsis on the wrong sylLABle. 4:15–26.

Frost, R. (2012). A universal approach to modeling visual word recognition and reading: Not only possible, but also inevitable. *Behavioral and Brain Sciences*, 35(05):310–329.

Fudge, E. (1984). *English word-stress*. Routledge.

Ganushchak, L. and Schiller, N. (2008). Motivation and semantic context affect brain error-monitoring activity: an event-related brain potentials study. *Neuroimage*, 39(1):395–405.

Ganushchak, L. and Schiller, N. (2009). Speaking one's second language under time pressure: An ERP study on verbal self-monitoring in German–Dutch bilinguals. *Psychophysiology*, 46(2):410–419.

Garrett, M. (1976a). The analysis of sentence production. *Psychology of Learning and Motivation:*, 9:133.

Garrett, M. (1976b). Syntactic processes in sentence production. *New Approaches to Language Mechanisms*, 30:231–256.

Garrett, M. F. (1992). Lexical retrieval processes: Semantic field effects. *Frames, fields and contrasts: New essays in semantic and lexical organization*, pages 377–395.

Gerlach, B. and Grijzenhout, J. (2000). Clitics in phonology, morphology and syntax. *Linguistik Aktuell*, 36.

Giegerich, H. (2004). Compound or phrase? english noun-plus-noun constructions and the stress criterion. *English Language and Linguistics*, 8(01):1–24.

Göksel, A. (2009). Compounds in Turkish. *Lingue e Linguaggio*, 8(2):213–236.

Gollan, T. and Acenas, L.-A. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and Tagalog-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1):246.

Gollan, T., Montoya, R., Fennema-Notestine, C., and Morris, S. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, 33(7):1220–1234.

Gollan, T. and Silverberg, N. (2001). Tip-of-the-tongue states in Hebrew-English bilinguals. *Bilingualism: language and cognition*, 4(01):63–83.

Grainger, J. and Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes*, 13(6):601–623.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(02):67–81.

Green, D. and Abutalebi, J. (2008). Understanding the link between bilingual aphasia and language control. *Journal of Neurolinguistics*, 21(6):558–576.

Green, D. and Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5):515–530.

Griffin, Z. and Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, 38(3):313–338.

Grosjean, F., Bhatia, T., and Ritchie, W. (2004). Studying bilinguals: Methodological and conceptual issues. *The Handbook of Bilingualism*, pages 32–63.

Gut, U. (2003). Prosody in second language speech production: The role of the native language. *Fremdsprachen Lehren und Lernen*, 32:133–152.

Halle, M. and Vergnaud, J. (1978). *Metrical structures in phonology*. Cambridge, MA.

Halle, M. and Vergnaud, J. (1987). *An essay on stress*. MIT press.

Hannagan, T., Magnuson, J., and Grainger, J. (2013). Spoken word recognition without a trace. *Frontiers in Psychology*, 4:563.

Hannahs, S. (1995). The phonological word in French. *Linguistics*, 33(6):1125–1144.

Hanulová, J., Davidson, D., and Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes*, 26(7):902–934.

Harley, T. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8(3):191–219.

Hatch, E. (1983). *Psycholinguistics: A Second Language Perspective.* Longman Higher Education.

Haugen, E. (1956). The syllable in linguistic description. In M. Halle, H. L. and McClean, H., editors, *For Roman Jakobson*, pages 213–221. Mouton The Hague.

Hayes, B. (1980). *A metrical theory of stress rules*. PhD thesis, Massachusetts Institute of Technology.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry*, 15(1):33–74.

Hayes, B. (1989). The prosodic hierarchy in meter. In Kiparsky, P. and Youmans, G., editors, *Rhythm and Meter: Phonetics and Phonology*, volume 1, pages 201–260. Academic Press.

Hayes, B. and Lahiri, A. (1991). Bengali intonational phonology. *Natural Language & Linguistic Theory*, 9(1):47–96.

Hirsh-Pasek, K., Nelson, D., Jusczyk, P., Cassidy, K., Druss, B., and Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26(3):269–286.

Hittmair-Delazer, M., Semenza, C., and Denes, G. (1994). Concepts and facts in calculation. *Brain*, 117(4):715–728.

Hockett, C. (1955). *A manual of phonology*. Number 11. Waverly Press.

Holroyd, C. and Yeung, N. (2003). Alcohol and error processing. *Trends in Neurosciences*, 26(8):402–404.

Hoshino, N. and Kroll, J. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, 106(1):501–511.

Hoshino, N. and Thierry, G. (2011). Language selection in bilingual word production: electrophysiological evidence for cross-language competition. *Brain Research*, 1371:100–109.

Humphreys, G., Evett, L., and Quinlan, P. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, 22(4):517–560.

Hunyadi, L. (2010). Cognitive grouping and recursion in prosody. *Recursion and Human Language*, 104(18):343.

Inhoff, A. W., Briihl, D., and Schwartz, J. (1996). Compound word effects differ in reading, on-line naming, and delayed naming tasks. *Memory & Cognition*, 24(4):466–476.

Inkelas, S. (1991). *Prosodic constituency in the lexicon*. Garland Publishing.

Ionin, T. and Wexler, K. (2002). Why is 'is' easier than '-s'?: acquisition of tense/agreement morphology by child second language learners of english. *Second Language Research*, 18(2):95–136.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. *VerbPparticle Explorations*, 1:67.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception*, 26(3):195–204.

Jacobs, C. and Dell, G. (2014). 'hotdog', not 'hot' 'dog': the phonological planning of compound words. *Language, Cognition and Neuroscience*, 29(4):512–523.

Jacquemot, C. and Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, 10(11):480–486.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 1(2):130.

James, W. (1890). *The principles of psychology*, volume 1. New York: Holt.

Janssen, N., Bi, Y., and Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes*, 23(7-8):1191–1223.

Janssen, N., Pajtas, P. E., and Caramazza, A. (2014). Task influences on the production and comprehension of compound words. *Memory & Cognition*, 42(5):780–793.

Jenkins, N. (2013). IETLS: for educational institutions, governments, professional bodies and commercial organisations.

Jescheniak, J. D. and Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4):824.

Jescheniak, J. D., Schriefers, H., and Hantsch, A. (2003). Utterance format effects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2):441–454.

Jones, D. M., Macken, W. J., and Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3):656.

Jónsson, G. (1954). *Eddukvaeði.(Saemundar-Edda)*, volume 1. Íslendingasagnaútgáfan.

Jordan, M. (1990). *Motor learning and the degrees of freedom problem.* Lawrence Erlbaum Associates, Inc.

Juhasz, B., Starr, M., Inhoff, A., and Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, 94:223.

Kabak, B. and Vogel, I. (2001). The phonological word and stress assignment in Turkish. *Phonology*, 18(03):315–360.

Kaisse, E. (1983). The syntax of auxiliary reduction in english. *Language*, pages 93–122.

Kanerva, J. (1990). Focusing on phonological phrases in Chichewa. In *The Phonology-Syntax Connection*, pages 145–162. University of Chicago Press.

Kemmer, S. (1993). *The Middle Voice*, volume 23. John Benjamins Publishing.

Kiparsky, P. (1982). From cyclic phonology to lexical phonology. *The Structure of Phonological Representations*, 1:131–175.

Klapp, S. and Erwin, C. (1976). Relation between programming time and duration of the response being programmed. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4):591.

Kolers, P. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior*, 2(4):291–300.

Konopka, A. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, 66(1):143–162.

Kovelman, I., Baker, S., and Petitto, L.-A. (2008). Bilingual and monolingual brains compared: a functional magnetic resonance imaging investigation of syntactic processing and a possible "neural signature" of bilingualism. *Journal of Cognitive Neuroscience*, 20(1):153–169.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, 35(2):162–179.

Kroll, J. (1993). Accessing conceptual representations for words in a second language. In *The Bilingual Lexicon*, volume 6, pages 53–77. Schreuder, R and Weltens, B.

Kroll, J., Bobb, S., and Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9(02):119–135.

Kroll, J., Tokowicz, N., et al. (2005). Models of bilingual representation and processing. In *Handbook of Bilingualism: Psycholinguistic Approaches*, pages 531–553. Oxford University Press New York.

Kussmaul, A. (1877). Disturbances of speech. In Ziemssen, H. v., editor, *Cyclopaedia of the Practice of Medicine*, volume 14, pages 581–875. New York: William Wood.

Labov, W. (1969). Contraction, deletion, and inherent variability of the english copula. *Language*, pages 715–762.

Ladd, R. (1984). English compound stress. In Gibbon, D. and Richter, H., editors, *Intonation, accent and rhythm*, pages 253–266. Berlin, de Gruyter.

Ladd, R. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook*, 3(311):40.

Ladd, R. (1996). *Intonational Phonology*. Cambridge University Press.

Lahiri, A. and Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38(3):245–294.

Lahiri, A. and Plank, F. (2010). Phonological phrasing in Germanic: the judgement of history, confirmed through experiment. *Transactions of the Philological Society*, 108(3):370–398.

Lakoff, G. (1970). *Irregularity in Syntax*. Holt, Rinehart and Winston, New York.

Lange, V. (2014). Inter-subject variability modulates phonological advance planning in the production of adjective-noun phrases. *Frontiers in Psychology*, 31(5):1–12.

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent l2 end-state grammar. *Second language research*, 14(4):359–375.

Lehiste, I. (1970). *Suprasegmentals*. M.I.T. Press.

Levelt, W. (1989). *Speaking*. MIT Press.

Levelt, W. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1):1–22.

Levelt, W. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, 98(23):13464–13471.

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–75.

Li, A. and Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm. *Studies in Second Language Acquisition*, 36(02):223–255.

Libben, G., Gibson, M., Yoon, Y., and Sandra, D. (1997). Semantic transparency and compound fracture. *CLASNET Working Papers*, 9:1–13.

Liberman, M. (1975). *The intonational system of English*. PhD thesis, Massachusetts Institute of Technology.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2):249–336.

Liberman, M. and Sproat, R. (1992). The stress and structure of modified noun phrases in english. *Lexical Matters*, pages 131–181.

Lichtheim, L. (1885). Sur l'aphasie. *Revue Philosophique de la France Et de l'Etranger*, 19(588).

Lipski, J. (1978). Code-switching and the problem of bilingual competence. In *Aspects of Bilingualism*, volume 250, page 264. Hornbeam Press Columbia, SC.

Loftus, E. and Yuille, J. (1984). Departures from reality in human perception and memory. *Memory Consolidation: Psychobiology of Cognition*, pages 163–184.

Luck, S. and Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–281.

Lukatela, G. and Turvey, M. (2000). Do spelling variations affect associative and phonological priming by pseudohomophones? *Perception & Psychophysics*, 62(1):196–217.

MacKay, D. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, 89(5):483.

MacKay, D. (1987). Constraints on theories of sequencing and timing in language perception and production. In *Language perception and production: Relationships between listening, speaking, reading and writing*, volume 3, page 407. Academic Press London.

MacKay, D. and Miller, M. (1994). Semantic blindness: Repeated concepts are difficult to encode and recall under time pressure. *Psychological Science*, 5(1):52–55.

Macnamara, J. (1967). The linguistic independence of bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 6(5):729–736.

Mahon, B., Costa, A., Peterson, R., Vargas, K., and Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3):503.

Mandel, D., Nelson, D. G., and Jusczyk, P. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development*, 11(2):181–196.

Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. C. H. Beck'sche Verlagsbuchhandlung.

Marian, V., Blumenfeld, H., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (leap-q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4):940–967.

Marslen-Wilson, W. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1):29–63.

Martin, R., Crowther, J., Knight, M., Tamborello, II, F., and Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, 116(2):177–192.

Martin, R. and Freedman, M. (2001). Short-term retention of lexical-semantic representations: Implications for speech production. *Memory*, 9(4-6):261–280.

McCarthy, J. (1979). On stress and syllabification. *Linguistic inquiry*, 10(3):443–465.

McCormack, P. (1974). Bilingual linguistic memory: Independence or interdependence; two stores or one. In Carey, S., editor, *Bilingualism, biculturalism and education.* Academic Press New York.

McCormack, P. (1977). Bilingual linguistic memory: The independence-interdependence issue revisited. In Hornsby, P., editor, *Bilingualism: Psychological, social, and educational implications*, pages 57–66. Academic Press New York.

McMahon, A. (1967). Phonemes and phonemic units of Cora (Mexico). *International Journal of American Linguistics*, 33:128–134.

McTurk, R. (2008). *A companion to Old Norse-Icelandic literature and culture*. John Wiley & Sons.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20(01):83–108.

Meisel, J. (1997). The acquisition of the syntax of negation in french and german: Contrasting first and second language development. *Second language research*, 13(3):227–263.

Meringer, R. and Mayer, K. (1895). Versprechen und verlesen: eine psychologish-linguistische studie. *Stuttgart: Göschensche Verlagsbuschhandlung*.

Meyer, A. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, 29(5):524–545.

Meyer, A. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, 30(1):69–89.

Meyer, A. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, 42(1):181–211.

Meyer, A. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, 35:477–496.

Meyer, A., Roelofs, A., and Levelt, W. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, 48(1):131–147.

Meyer, A. and Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(6):1146.

Milroy, J. and Milroy, L. (1985). Authority in language. *Investigating Language Prescription and Standardisation. London/New York*.

Miozzo, M. and Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4):907.

Nakayama, M., Sears, C., Hino, Y., and Lupker, S. (2012). Cross-script phonological priming for Japanese-English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes*, 27(10):1563–1583.

Nespor, M. and Ralli, A. (1996). Morphology-phonology interface: Phonological domains in greek compounds. *The Linguistic Review*, 13(3-4):357–382.

Nespor, M. and Sandler, W. (1999). Prosody in Israeli sign language. *Language and Speech*, 42(2-3):143–176.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*, volume 28. Dordrecht: Foris.

Nespor, M. and Vogel, I. (2007). *Prosodic phonology*, volume 28. Walter de Gruyter, 2 edition.

Newman, J. (1951). The phonetic aspect of Joshua Steele's system of prosody. *Communications Monographs*, 18(4):279–287.

Nicoladis, E. and Yin, H. (2002). The role of frequency in acquisition of English and Chinese compounds by bilingual children. In *Proceedings of the Boston University Conference on Language Development*, pages 441–452.

Nikolou, K. (2011). The phonological word in Greek and its comparison with Arabic: Findings and applications. *Journal of Greek Linguistics*, 11(1):122–124.

Oldfield, R. and Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4):273–281.

Oomen, C. C. and Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, 30(2):163–184.

Oppermann, F., Jescheniak, J. D., and Schriefers, H. (2010). Phonological advance planning in sentence production. *Journal of Memory and Language*, 63(4):526–540.

Ottósson, K. (2008). The Old Nordic middle voice in the pre-literary period. In Folke Josephson, I. S., editor, *Interdependence of Diachronic and Synchronic Analyses*, volume 103, pages 185–220. John Benjamins Publishing.

Page, M., Madge, A., Cumming, N., and Norris, D. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language*, 56(1):49–64.

Paivio, A. (2014). Bilingual dual coding theory and memory. In *Foundations of Bilingual Memory*, pages 41–62. Springer.

Paivio, A. and Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 34(4):388.

Palij, M. and Aaronson, D. (1992). The role of language background in cognitive processing. *Advances in Psychology*, 83:63–87.

Palmer, C. and Kelly, M. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language*, 31(4):525–542.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, 116(2):220.

Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Halle: Niemeyer.

Peperkamp, S. (1997). *Prosodic words*, volume 34. Holland Academic Graphics: The Hague.

Peperkamp, S., Dupoux, E., and Sebastián-Gallés, N. (1999). Perception of stress by french, spanish, and bilingual subjects. In *Eurospeech*, pages 2683–2686.

Perales, S. and Liceras, J. (2010). Looking for universals in the acquisition of L2 Spanish object clitics. In *New directions in Language Acquisition.*, pages 419–51. Cambridge Scholars Publishing: UK.

Perfetti, C. and Tan, L. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(1):101.

Pexman, P., Lupker, S., and Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1):139.

Pexman, P., Lupker, S., and Reggin, L. (2002). Phonological effects in visual word recognition: investigating the impact of feedback activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3):572.

Pinker, S. and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, 95(2):201–236.

Plag, I. (2006). The variability of compound stress in English: structural, semantic, and analogical factors. *English Language and Linguistics*, 10(01):143–172.

Plag, I., Kunter, G., Braun, M., and Lappe, S. (2007). Testing hypotheses about compound stress assignment in English: a corpus-based investigation. *Corpus Linguistics and Linguistic Theory*, 3(2):199–232.

Poarch, G. and van Hell, J. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, 113(4):535–551.

Poeppel, D. and Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, 92(1):1–12.

Pollatsek, A., Hyönä, J., and Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2):820.

Potter, M. C., So, K.-F., Von Eckardt, B., and Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23(1):23–38.

Poulisse, N. and Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15(1):36–57.

Poulisse, N., de Groot, A., and Kroll, J. (1997). Language production in bilinguals. *Tutorials in Bilingualism: Psycholinguistic Perspectives*, 201:224.

Prince, A. and Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.

Probert, P. (2006). *Ancient Greek Accentuation: Synchronic Patterns, Frequency Effects, and Prehistory: Synchronic Patterns, Frequency Effects, and Prehistory*. OUP Oxford.

Qu, Q., Damian, M., and Li, X. (2015). Phonology contributes to writing: evidence from a masked priming task. *Language, Cognition and Neuroscience*, pages 1–14.

R Core Team (2015). R: A language and environment for statistical computing.

Rastle, K. and Brysbaert, M. (2006). Masked phonological priming effects in english: Are they real? do they matter? *Cognitive Psychology*, 53(2):97–145.

Rastle, K. and Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2):307–314.

Reetz, H. (2010). "voicekeyhilary.pl". [Perl Computer program].

Revithiadou, A. (2011). The phonological word. In van Oostendorp, M., editor, *The Blackwell Companion to Phonology: Suprasegmental and Prosodic Phonology*, volume 2, pages 1204–1227. Wiley-Blackwell.

Revithiadou, A. and Spyropoulos, V. (2008). Greek object clitic pronouns: A typological survey of their grammatical properties. *STUF-Sprachtypologie und Universalienforschung*, 61(1/2008):39–53.

Roberts, P. M., Garcia, L. J., Desrochers, A., and Hernandez, D. (2002). English performance of proficient bilingual adults on the boston naming test. *Aphasiology*, 16(4-6):635–645.

Rodriguez-Fornells, A., Van Der Lugt, A., Rotte, M., Britti, B., Heinze, H.-J., and Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, 17(3):422–433.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, pages 107–142.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, 47(1):59–87.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64(3):249–284.

Roelofs, A. (2003). Shared phonological encoding processes and representations of languages in bilingual speakers. *Language and Cognitive Processes*, 18(2):175–204.

Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6):1580–1598.

Roelofs, A. and Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4):922.

Roelofs, A. and Verhoef, K. (2006). Modeling the control of phonological encoding in bilingual speakers. *Bilingualism: Language and Cognition*, 9(02):167.

Rouder, J., Morey, R., Morey, C., and Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, 18(2):324–330.

Runnqvist, E. and Costa, A. (2012). Is retrieval-induced forgetting behind the bilingual disadvantage in word production? *Bilingualism: Language and Cognition*, 15(02):365–377.

Sandler, W. (1999). Cliticization and prosodic words in a sign language. *Studies on the Phonological Word*, 174:223.

Saran, F. (1907). *Deutsche Verslehre*, volume 3. CH Beck'sche verlagsbuchhandlung. O. Beck.

Scheer, T. (2011). *A guide to morphosyntax-phonology interface theories: how extra-phonological information is treated in phonology since Trubetzkoy's Grenzsignale*. Walter de Gruyter.

Schiller, N. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, 39(3):484–507.

Schiller, N. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2):512.

Schiller, N. and Costa, A. (2006). Activation of segments, not syllables, during phonological encoding in speech production. *The Mental Lexicon*, 1(2):231–250.

Schiller, N., Costa, A., and Colomé, A. (2002). Phonological encoding of single words: In search of the lost syllable. *Papers in Laboratory Phonology*, 7:35–59.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). E-prime computer software and manual. *Pittsburgh, PA: Psychology Software Tools Inc*.

Schnur, T. T. (2011). Phonological planning during sentence production: beyond the verb. *Frontiers in Psychology*, 2.

Schnur, T. T., Costa, A., and Caramazza, A. (2006). Planning at the Phonological Level during Sentence Production. *Journal of Psycholinguistic Research*, 35(2):189–213.

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition*, 45(1):33–54.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4):841.

Schriefers, H., Meyer, A. S., and Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language*, 29(1):86–102.

Schriefers, H. and Teruel, E. (1999). The production of noun phrases: A cross-linguistic comparison of french and german. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, pages 637–642.

Schütze, C. T. and Ferreira, V. S. (2007). What should we do with our speech error corpora? notes from the panel discussion.

Selkirk, E. (1972). *The Phrase Phonology of English and French*. PhD thesis, MIT.

Selkirk, E. (1980a). *On prosodic structure and its relation to syntactic structure*. Indiana University Linguistics Club.

Selkirk, E. (1980b). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11(3):563–605.

Selkirk, E. (1984). *Phonology and syntax*, volume 10 of *Current studies in linguistics*. MIT Press.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology*, 3(01):371–405.

Selkirk, E. (2011). The syntax-phonology interface. In John Goldsmith, J. R. and Yu, A. C., editors, *The Handbook of Phonological Theory*, pages 435–484. Wiley.

Semenza, C., Luzzatti, C., and Carabelli, S. (1997). Morphological representation of compound nouns: A study on Italian aphasic patients. *Journal of Neurolinguistics*, 10(1):33–43.

Sen, S. (2015). Abolishing english in schools: implications for higher education in west bengal. *Inter-Asia Cultural Studies*, 16(2):269–281.

Shao, Z., Roelofs, A., Martin, R., and Meyer, A. (2013). Selective inhibition and naming performance in semantic blocking, picture–word interference, and color-word stroop tasks. *Manuscript submitted for publication*.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In *Sentence Processing: Psycholinguistic Studies presented to Merrill Garrett*, pages 295–342.

Shi, L.-F. (2015). English word frequency and recognition in bilinguals: Inter-corpus comparison and error analysis. *International journal of audiology*, 54(10):674–681.

Sholl, A., Sankaranarayanan, A., and Kroll, J. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science*, 6(1):45–49.

Shukla, M., White, K., and Aslin, R. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, 108(15):6038–6043.

Sievers, E. (1901). *Grundzüge der phonetik*. Breitkopf & Härtel.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). Tobi: a standard for labeling english prosody. In *The Second International Conference on Spoken Language Processing, ICSLP 1992, Banff, Alberta, Canada, October 13-16, 1992*.

Simmonds, A., Wise, R., and Leech, R. (2011). Two tongues, one brain: imaging bilingual speech production. *Frontiers in Psychology*, 2:1–13.

Smith, M. and Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, 73(3):205–246.

Snodgrass, J. and Yuditsky, T. (1996). Naming times for the snodgrass and vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, 28(4):516–536.

Sproat, R. (1992). *Morphology and computation*. MIT Press.

Steele, J. (1779). *Prosodia Rationalis: Or, An Essay Towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols*. J. Nichols, sold by T. Payne.

Stemberger, J. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua*, 56(3):235–259.

Stemberger, J. (1985). An interactive activation model of language production. *Progress in the Psychology of Language*, 1:143–186.

Sternberg, S., Monsell, S., Knoll, R., and Wright, C. (1978). The Latency and Duration of Rapid Movement Sequences: Comparisons of Speech and Typewriting. In *Information Processing in Motor Control and Learning*, pages 117–152. Academic Press.

Storkel, H. (2013). A corpus of consonant–vowel–consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods*, 45(4):1159–1167.

Sturlasson, S. (1875). *Edda: Skáldskaparmál*. Viking Society for Northern Research, University College.

Sturtevant, E. (1917). *Linguistic Change: An Introduction to the Historical Study of Language*, volume 60. University of Chicago Press.

Sweet, H. (1886). *Elementarbuch des gesprochenen Englisch (Grammatik, Texte und Glossar)*. Clarendon Press.

Taylor, J., Rastle, K., and Davis, M. (2014). Interpreting response time effects in functional imaging studies. *NeuroImage*, 99(C):419–433.

Ueyama, M. (2000). *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese*. PhD thesis, University of California Los Angeles.

Unsworth, N. and Engle, R. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1):104.

Vagias, W. M. (2006). Likert-type scale response anchors. *Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University*.

van der Hulst, H. (2010). *Recursion and Human Language*, volume 104. Walter de Gruyter.

Van Oostendorp, M. (2002). The phonological and morphological status of the prosodic word adjunct. *Linguistische Berichte Sonderheft*, 11:209–236.

van Ostade, I. (2008). *Grammars, grammarians and grammar-writing in eighteenth-century England*, volume 59. Walter de Gruyter.

Vigário, M. (2010). Prosodic structure between the prosodic word and the phonological phrase: Recursive nodes or an independent domain? *The Linguistic Review*, 27(4).

Vogel, I. (1994). Phonological interfaces in Italian. *Proceedings of Linguistic Studies on Romance Languages XXIII*.

Vogel, I. (2009). The status of the clitic group. *Phonological Domains: Universals and Deviations*, 16:15.

Wagner, V., Jescheniak, J., and Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2):423–440.

Warren, R. et al. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917):392–393.

Watson, I. (1991). Phonological processing in two languages. *Language Processing in Bilingual Children*, pages 25–48.

Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Cohn.

Whaley, C. (1978). Word–nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17(2):143–154.

Wheat, K., Cornelissen, P., Frost, S., and Hansen, P. (2010). During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: evidence from magnetoencephalography. *The Journal of Neuroscience*, 30(15):5229–5233.

Wheeldon, L. and Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37(3):356–381.

Wheeldon, L. and Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical word. *Cognition*, 85(2):B31–B41.

Wheeldon, L. and Morgan, J. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes*, 17(5):503–535.

Wheeldon, L. R. and Levelt, W. J. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34(3):311–334.

Winter, B. (2013). Linear models and linear mixed effects models in r with linguistic applications.

Woutersen, M. (1997). *Bilingual word perception*. PhD thesis.

Zwicky, A. M. (1977). *On clitics*. Indiana University Linguistics Club.